BAYESIAN INVERSE MODELING: ALGORITHMS AND APPLICATIONS IN GROUND WATER

A Dissertation
Presented to
The Academic Faculty

by

Yue Zhao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
May 2020

BAYESIAN INVERSE MODELING: ALGORITHMS AND APPLICATIONS IN GROUND WATER

Approved by:

Dr. Jian Luo
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Jingfeng Wang
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Aris P. Georgakakos
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Emanuele Di Lorenzo
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. Francesco Fedele
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: 03/19/2020

To my parents

ACKNOWLEDGEMENTS

There many people who deserve recognition and thanks for their time and efforts in leading to the completion of this thesis. I would like to express my sincere gratitude to my advisor, Dr. Jian Luo for his support, guidance, encouragement, and equally importantly, an environment of high degree of autonomy to explore novel ideas and topics, and for conducting quality research. This work would not have been possible without his help.

I would also like to acknowledge my committee members: Dr. Aris P. Georgakakos, Dr. Francesco Fedele, Dr. Jingfeng Wang and Dr. Emanuele Di Lorenzo for their constructive comments, suggestions and salient advice to help complete this thesis.

I acknowledge our laboratory group members of enriching my life: Quan Guo, Jinge Huo, Ming Liu, Saubhagya Singh Rathore, Yuening Tang, Minjae Kim, and former lab members Chunhui Lu and Rulan Gong. I also acknowledge the members of the Water Resources Engineering and the friends I have made in Atlanta; special thanks to Yuzhi Guo, Shihyu Huang, Martin Jang, Longde Jin, Fangzhou Liu, Xiaofeng Liu, Mahdi Roozbahani, Yao Tang, Bo Wang, Tongtong Xu.

I would also like to extend my gratitude to the following people who have helped in various ways: to Dr. Deqiang Mao for him providing me field data; to Fangzhou Liu for friendship, mind-stimulating discussions; to Ming Liu for the relentless encouragement in pursuing knowledge in computational science and engineering; to Saubhagya Singh Rathore for the reminders whenever my priorities become sidetracked.

Finally, I would like to express my wholehearted gratitude to my parents for their unconditional love, timely encouragement and endless patience not only to complete this thesis but also to take a journey of my life this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$\mathbf{y}$	Available measurement.
f	Forward model.
ϵ	Error involved in the measurement equation.
$\mathbf{s}$	Interested random variable in the inverse problem.
$\mathcal{R}$	Regularization term.
$\mathbf{T}$	Tikhonov regularization matrix.
υ	Parameter for regularization term.
$\mathbf{R}$	Covariance matrix of the error term ϵ.
σ^2	Variance of the error term ϵ when it is assumed as i.i.d..
ℓ	Loss function derived from posterior distribution.
$\mathbf{C}$	Covariance matrix of $\mathbf{s}$.
$\mathbf{V}$	All the principal component.
$\mathbf{v}$	One principal component.
$\mathbf{D}$	Diagonal matrix storing all eigenvalues after eigen decomposition.
$\mathbf{V}_k$	Truncated, the first k principal components.
$\boldsymbol{\mu}_\mathbf{s}$	Mean vector of the interested random variable.
$\mathbf{a}$	Reduced random variable, principal component coefficient.
$\boldsymbol{\Sigma}_a$	Covariance matrix of $\mathbf{a}$.
$\mathbf{A}$	Hessian matrix of loss function ℓ.
$\mathbf{H}_\mathbf{a}$	Jacobian matrix of principal component coefficient $\mathbf{a}$.
$\boldsymbol{\beta}$	Drift coefficient.
$\mathbf{H}_{\boldsymbol{\beta}}$	Jacobian matrix of drift coefficient $\boldsymbol{\beta}$.

$\bar{\mathbf{H}}'_{\mathbf{a}}$	Jacobian matrix of principal component coefficient **a** in quasi-linear approach.
$\bar{\mathbf{H}}'_{\boldsymbol{\beta}}$	Jacobian matrix of drift coefficient **β** in quasi-linear approach.
$\bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}}$	Jacobian matrix of principal component coefficient **a** and drift coefficient **β** in quasi-linear approach.
$\bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}}$	Jacobian matrix of principal component coefficient **a** and drift coefficient **β** for logarithmic storage coefficient in quasi-linear approach.
K	Hydraulic conductivity.
U	Upscaling matrix.
$\mathbf{s}_2$	Upscaled interested random variable.
$\boldsymbol{\epsilon}_2$	Error in measurement equation for upscaled forward model.
δ	A small delta parameter for the matrix-free approach.
X	Drift of mean.
$\boldsymbol{\xi}$	Intermediate variable in the quasi-linear approach.
$\boldsymbol{\Lambda}$	Matrix in cokriging inverse approach.
$\mathbf{X}'$	Upscaled drift of mean.
$\mathbf{V}'_{\mathrm{k}}$	Upscaled, truncated principal components.
q	Logarithmic storage coefficient.
Y	Drift of mean of **q**.
$\mathbf{W}_{\mathrm{k}}$	Truncated, the first k principal components of **q**.
w	Principal component of **q**.
$\boldsymbol{\beta}_{\mathbf{q}}$	Drift coefficient of **q**.
$\mathbf{a}_{\mathbf{q}}$	Reduced random variable, principal component coefficient of **q**.
$\mathbf{Y}'$	Upscaled drift of mean of **q**.
$\mathbf{W}'_{\mathrm{k}}$	Upscaled, truncated principal components of **q**.
$\boldsymbol{\epsilon}_{2,2}$	Error involved in the measurement equation given $\mathrm{f}(\mathbf{s}, \mathbf{q})$.

$\mathbf{R_{2,2}}$	Covariance matrix of $\boldsymbol{\epsilon}_{2,2}$.
$\sigma_{2,2}^2$	Variance of the error term $\boldsymbol{\epsilon}_{2,2}$ when it is assumed as i.i.d..

SUMMARY

Inverse estimation of spatially-correlated parameter fields is essential in a variety of scientific disciplines, including hydrology, geology, geophysics, natural resources, environmental sciences and engineering, etc. Classical stochastic sampling approach, such as Markov Chain Monte Carlo (MCMC), and optimization approach, such as geostatistical approach (GA), are capable of solving inverse problems having a modest size of 10^3-10^4 unknowns with fast forward model evaluation. However, the computational cost becomes unaffordable for large-dimensional inverse problems with 10^6 unknowns on fine-resolution parameter fields. In this thesis, we develop new, efficient and accurate algorithms for both stochastic sampling approach and geostatistical approach within a Bayesian framework and test the efficacy using synthetic and field data for inverse estimation of large-dimensional hydrogeological parameter fields, including hydraulic conductivity and specific storage.

For stochastic sampling, the stochastic Newton algorithm is employed to propose effective samples to speed up convergence, taking advantage of the local structure of posterior distribution. A quasi-Newton algorithm is incorporated into stochastic Newton to accelerate the sampling process by reducing the number of forward model evaluations. For the application of hydraulic tomography, an upscaling approach is applied to reduce the computational time of each forward model evaluation while still preserves the inverse accuracy. The proposed approach is validated via numerical experiments using synthetic datasets.

For geostatistical approach, we reformulate the quasi-linear geostatistical approach onto principal component coefficients that significantly reduces the number of forward

model runs, and benefits post-solution uncertainty analysis and generation of conditional realizations. This approach yields a more scalable normal equation system that is particularly advantageous for inverse problems with a large number of measurements, and therefore will potentially become popular in ‘big data’ era. The computational efficiency is improved further by combining an upscaling approach, which significantly lowers the running time of forward model evaluations. The proposed approach is successfully applied to both synthetic and field cases.

CHAPTER 1. INTRODUCTION

In groundwater hydrology, aquifer simulations contribute to a better understanding of groundwater flow and transport behavior that can be exploited to guide in-situ remediation or groundwater resources management. An aquifer simulation requires characterization of hydrogeologic properties such as hydraulic conductivity and specific storage coefficient, which are inherently heterogeneous in space. In hydrogeology, local-scale estimates of hydraulic conductivities can be obtained by laboratory measurements of core samples [*Klute and Dinauer,* 1986a]. However, high costs in drilling wells on highly discretized grids hamper a broader application of this approach in field practice. Therefore, characterization of hydraulic conductivity is usually achieved by inference from indirect measurements such as slug tests, pumping tests, tracer tests, etc. [*Cardiff et al.,* 2009; *Cirpka et al.,* 2007; *Fienen et al.,* 2006]. The inference approach is also called as an inverse approach as it derives input parameters from measurable outputs. This thesis mainly focuses on inverse modeling of large-dimensional hydrogeologic parameter fields such as hydraulic conductivity, as it is usually considered as the major uncertainty source in aquifer simulations. Certainly, the proposed research approaches and computational algorithms can be readily extended to invert random fields of other physical properties in a variety of scientific disciplines, including hydrology, geology, geophysics, natural resources, agriculture, environmental sciences and engineering, etc.

An inverse problem can be conveniently formulated in a Bayesian framework [*Gavalas et al.,* 1976], which produces a posterior distribution constrained by the likelihood function that encodes the data information and the prior that described the most

favorable patterns of the underlying field. Both ensemble-based and non-ensemble-based computational approaches are proposed for solving Bayesian inverse problems. Representative members from each category are Markov chain Monte Carlo (MCMC) [*Geman and Geman,* 1984a; *Hastings,* 1970] and the geostatistical approach (GA) [*Hoeksema and Kitanidis,* 1984; *Kitanidis,* 1995]. Classical inverse approaches are capable of estimating parameter fields such as hydraulic conductivities with a modest size of 10^3-10^4 unknowns based on hydrogeologic testing measurements. However, if the dimension increases to 10^6, the computational cost becomes prohibitive. Continued efforts have been dedicated to improve the computational efficiency of these approaches in the past decades [*Cui et al.,* 2011; *Cui et al.,* 2014; *Efendiev et al.,* 2005; *Lin et al.,* 2017; *Liu and Kitanidis,* 2011]. However, due to the growing demand of high-resolution characterization of the field and the complexity of forward models, the computational cost of solving a high-dimensional Bayesian inverse problem with millions of unknowns remains prohibitive.

The main goal of this thesis is to improve the efficiency of MCMC algorithm and geostatistical approach for large-scale and high-dimensional Bayesian inverse problems in hydrogeology. For MCMC, we propose to incorporate a quasi-Newton approach, the Broyden method, and an upscaling approach into a stochastic Newton framework to reduce the computational cost of running forward models during the sampling process. For the geostatistical approach, we propose a reformulated framework of the principal component geostatistical approach (PCGA) [*Kitanidis and Lee,* 2014], which is a quasi-linear approach, and an upscaling approach to improve its scalability, named as the upscaling PCGA (UPCGA). The proposed algorithms are validated by various numerical

experiments: transient pumping data inversions, hydraulic tomography problems and joint inversions of hydraulic conductivity and specific storage using field data.

1.1 Markov Chain Monte Carlo (MCMC)

MCMC is an ensemble-based approach that achieves the inference by simulating a large number of samples from the posterior distribution. MCMC is of particular interest in practice because of its statistical rigorousness. However, MCMC does not initiate the simulation of posterior samples before the convergence. The excessive number of forward model evaluations by running Markov chains in this process limits its applications in large-scale hydrogeological inverse problems, especially when the forward models are computationally expensive. To achieve a more computationally efficient MCMC algorithm, several approaches are proposed:

(1) Principal component analysis (PCA) is assimilated into the stochastic Newton (SN) algorithm to achieve a reduced sampling space and a lower computational cost of the approximated Hessian.

(2) Employ a quasi-Newton (Broyden) algorithm to update Jacobian matrices between consecutive iterations. The quasi-Newton approach computes the Jacobian matrix approximation with only one forward model evaluation, and thus improves the sampling efficiency.

(3) By recognizing principal component coefficients as a group of latent variables, an upscaling approach is proposed so that the inversion can be achieved by only evaluating the forward model on a coarse resolution, which substantially reduces the computational overhead of each forward model run.

1.2 Geostatistical Approach (GA)

The geostatistical approach is an optimization-based approach for Bayesian inverse problems. Instead of characterizing the entire posterior distribution, the geostatistical approach only finds the maximum a posteriori (MAP) estimate, usually denoted as "the best estimates, and provides the associated uncertainty by its posterior covariance matrix [*Kitanidis,* 1996]. Therefore, compared with MCMC, the geostatistical approach requires a much smaller computational budget. However, for Bayesian inverse problems with highly parameterized conductivity fields, the computational overhead of the geostatistical approach is still a practical issue. This thesis mainly makes three contributions to the geostatistical approach:

1) PCGA is reformulated into a more scalable form, providing a normal equation system whose dimensionality is independent of the number of measurements. This approach is more advantageous for inversion of big dataset.
2) The reformulated approach also offers a more intuitive way of computing Jacobian matrices and conditional realizations.
3) An upscaling approach is incorporated into the reformulated geostatistical approach by utilizing the principal component coefficients as a group of latent variables. The new approach is named as UPCGA that accelerates the inversion and uncertainty quantification by merely running the coarse-grid forward model.

1.3 Organization of Thesis

The following of this thesis is composed of six chapters. In CHAPTER 2, a literature review regarding Bayesian inverse modeling, MCMC and geostatistical approach is

provided. In CHAPTER 3 and CHAPTER 4, we propose a new MCMC sampling approach, and validate it by several numerical experiments with synthetic data. In CHAPTER 5 and CHAPTER 6, we propose the reformulated geostatistical approach and the upscaling approach on top of it, which are validated by numerical experiments with synthetic data. In CHAPTER 7, the performance of UPCGA is scrutinized further by inverting real world data. In CHAPTER 8, we close the discussion by concluding remarks and potential future research direction.

CHAPTER 2. LITERATURE REVIEW

2.1 Formulation of Bayesian Inverse Problems

A measurement equation is fundamental to understand the concept of inverse problem, which can be generally described as:

$$\mathbf{y} = \mathrm{f}(\mathbf{s}) + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{n\times 1}$ represents the available data; **f** represents the forward model, which is mainly a physical process described by a group of governing equations and solved by numerical solvers; $\mathbf{s} \in \mathbb{R}^{m\times 1}$ represents the interested variables to be estimated by the inverse approach and $\boldsymbol{\epsilon}$ represents errors involved in the forward process, which can be from oversimplification of physical models, random measurement error, numerical solver error, etc. In hydrogeological applications, variable **s**, with hydraulic conductivity as the most representative member, is often recognized as the main source of uncertainty [*Carrera and Neuman,* 1986; *Kitanidis,* 1996, 1997], whose characterization is required for a trustworthy model simulation. In practice, one way to estimate hydraulic conductivity in a local scale is by core samples [*Klute and Dinauer,* 1986b]. However, the high expenses of well drilling prevents this method from broader applications. On the other hand, the acquisition of **y** data, which could be head data, tracer concentration data, or temperature data, is more convenient compared with that of **s**. Therefore, in a real world application, the problem is usually processed in an 'inverse' way: the estimate and the uncertainty of **s** are usually inferred by the available measurements **y**. The inverse problem can be viewed as an optimization problem as

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 \tag{2}$$

where $\hat{\mathbf{s}} \in \mathbb{R}^{m \times 1}$ represents the best estimate of **s** given the available data **y**. However, in real world applications, the high-dimensional unknown **s** and the sparse observational measurements **y** incurs extraordinary high parameter-to-data sensitivity and ill-posed properties to this optimization problem. To mitigate the inherent instability and non-uniqueness of inverse problems, regularizations are required.

The form of the regularization term $\mathcal{R}(\mathbf{s})$ is problem dependent. Tikhonov regularization is one of the most popular ones in inverse problems. A Tikhonov regularization term can be formulated as:

$$\mathcal{R}(\mathbf{s}) = \|\mathbf{T}\mathbf{s}\|^2 \tag{3}$$

where **T** is called Tikhonov matrix. The Tikhonov regularization term imposes smoothness onto the solution of the optimization problem in Equation (2) and thus compensate its non-uniqueness. As a smoothing operator, matrix **T** has various forms in different applications. After adding regularization, the inverse problem can be formulated as

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 + \upsilon \|\mathbf{T}\mathbf{s}\|^2 \tag{4}$$

whose loss function is

$$\ell(\mathbf{s}) = \frac{1}{2} \|\mathbf{y} - \mathrm{f}(\mathbf{s})\|^2 + \upsilon \|\mathbf{T}\mathbf{s}\|^2 \tag{5}$$

where υ is usually a tuning parameter that controls the relative weights between the fitting to the observational data and the smoothness of the solution **s**.

The formulation of the inverse problem can also be interpreted from a Bayesian view. According to Bayes' rule, the posterior distribution $\mathrm{p}(\mathbf{s}|\mathbf{y})$ of the unknown parameter **s** is proportional to the product of its prior distribution $\mathrm{p}(\mathbf{s})$ and the likelihood of the observational data $\mathrm{p}(\mathbf{y}|\mathbf{s})$

$$\mathrm{p}(\mathbf{s}|\mathbf{y}) \propto \mathrm{p}(\mathbf{y}|\mathbf{s})\mathrm{p}(\mathbf{s}) \tag{6}$$

The likelihood encodes information from observational data and the prior provides the former beliefs about **s**, both of which constrain the posterior distribution. For a zero-mean Gaussian error term, the likelihood function is

$$\mathrm{p}(\mathbf{y}|\mathbf{s}) \propto \exp\left(-\frac{1}{2}\big(\mathbf{y}-\mathrm{f}(\mathbf{s})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y}-\mathrm{f}(\mathbf{s})\big)\right) \tag{7}$$

where **R** represents the covariance matrix of the error **ϵ**. A zero-mean Gaussian prior can be described as

$$\mathrm{p}(\mathbf{s}) \propto \exp\left(-\frac{1}{2}\mathbf{s}^{\mathrm{T}}(\mathbf{T}^{\mathrm{T}}\mathbf{T})\mathbf{s}\right) \tag{8}$$

where $(\mathbf{T}^{\mathrm{T}}\mathbf{T})$ acts as the inverse of the covariance matrix. Following Equation (6), the posterior distribution is

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y}-f(\mathbf{s})\right)^{T}\mathbf{R}^{-1}\left(\mathbf{y}-f(\mathbf{s})\right)-\frac{1}{2}\mathbf{s}^{T}(\mathbf{T}^{T}\mathbf{T})\mathbf{s}\right) \quad (9)$$

For $\mathbf{R}=\sigma^2\mathbf{I}$, it can be simplified as

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\left(\frac{1}{2\sigma^2}\|\mathbf{y}-f(\mathbf{s})\|^2+\|\mathbf{T}\mathbf{s}\|^2\right)\right) \quad (10)$$

If $\sigma^2=\upsilon$, the posterior distribution is proportional to the exponential of a negative loss function

$$p(\mathbf{s}|\mathbf{y}) \propto \exp\left(-\ell(\mathbf{s})\right) \quad (11)$$

Therefore, to find the parameter **s** that minimizes the loss function in Equation (5) equals to find the **s** that maximizes the posterior in Equation (10), and the inverse problem given an explicit loss function is formulated as a Bayesian inverse problem.

The regularization terms in inverse problems, or the prior distribution in Bayesian inverse problems have various forms. *Chen et al.* [2008] solved an inverse problem by an optimization method, where the regularization term is selected as the second norm of the interested variable. In this case, the Tikhonov matrix is a scaled identity matrix $\mathbf{T}=\lambda\mathbf{I}$, and the regularization term is $\mathcal{R}(\mathbf{s})=\lambda^2\|\mathbf{s}\|^2$. From a Bayesian perspective, it equals to applying a zero-mean multi-Gaussian prior onto the variable **s** with covariance $\mathbf{C_s}=\frac{1}{\lambda}\mathbf{I}$. *Liu and Kitanidis* [2011] leverage the connection between optimization and Bayesian inverse problems to employ a sparse Tikhonov matrix to enforce the smoothness of the inversion result.

Another widely accepted form of the covariance matrix, especially in ground water research area, the one defined by geostatistical covariance function [*Matheron,* 1963]. This form has been applied to plenty research in hydrogeological applications [*Kitanidis,* 1995, 1996; *Lee et al.,* 2018; *Liu et al.,* 2002; *Liu et al.,* 2007; *Nowak and Cirpka,* 2006a].

By adding regularization terms or priors, a well-posed Bayesian inverse problem is formulated and has a unique solution. The extra advantage of using prior in Bayesian approach is to provide uncertainty quantifications for the best estimate. Various computational approaches has been developed to seek for solutions of Bayesian inverse problems, which can be divided into two categories: the ensemble-based approach and the non-ensemble-based approach.

2.2 Ensemble-Based Bayesian Inverse Approach

Ensemble-based Bayesian inverse approaches provide the best estimate and the uncertainty quantifications via samples or approximated samples from the posterior distribution. Among all the ensemble-based approaches, the most representative members are randomized maximum likelihood (RML), ensemble Kalman filter (EnKF) or smoother, and MCMC.

RML is firstly proposed by *Kitanidis and Vomvoris* [1983b], which can provide exact samples for Gaussian posterior distribution, or produces approximated samples from non-Gaussian posterior distribution. The essential idea of RML is to maximize several perturbed posterior distributions to form an ensemble of approximated samples. This approach has been successfully applied to inversion of heat tracer data and history matching problems [*Khaninezhad and Jafarpour,* 2014; *Zhang et al.,* 2014].

Ensemble Kalman filter is also in great favor for solving Bayesian inverse problems, which has a close connection to RML [*Reynolds et al.,* 2006]. However, EnKF does not requires any explicit computation of the parameters' gradient, instead, EnKF updates the parameter field via covariance matrix of simulated forward model output. EnKF has been making great progress in the past a few years. *Reynolds et al.* [2006] exploits the relationship between EnKF and Gauss-Newton optimization to invent a new EnKF approach that admits an explicit gradient computation but achieves a faster convergence. *Zhou et al.* [2012] proposes a normal-score transformation to preprocess the inverted field and the measurements to adapt them to the Gaussian assumption of EnKF. The proposed approach demonstrates its ability to handle hydraulic conductivity field with non-Gaussian statistics. *Zhang et al.* [2018] proposes an iterative local updating EnKF to deal with multi-modal posterior distribution.

Compared with the aforementioned approaches, which rely on Gaussian assumption or approximation of the posterior distribution, MCMC is a more statistically strict approach that simulates samples from the exact posterior. The samples simulated by running a stochastic process, which is a Markov chain in this specific case, are correlated samples from the posterior, if the Markov chain is constructed correctly. Therefore, by retaining samples every a few iterations on the chain, a description of the posterior can be achieved by their summary statistics.

Two most classical MCMC methods are the Metropolis algorithm and the Gibbs algorithm [*Geman and Geman,* 1984b; *Hastings,* 1970], which established the cornerstone of MCMC and its application in Bayesian hierarchical model or probabilistic graphical model. Metropolis algorithm can be regarded as a random walk algorithm that explores the

solution space randomly. These two classical MCMC methods have been applied to analysis of tracer test [*Fienen et al.,* 2006; *Zhao et al.,* 2018], which can be formulated and solved efficiently as a linear Bayesian inverse problem. In these two applications, even though the posterior distributions of the transfer function and the reaction rate coefficient are complicated, non-traditional distributions, the results of the inversion indicates that when the forward simulation is cheap, the Metropolis algorithm and the Gibbs algorithm is capable of solving Bayesian inverse problems.

Despite of its capability in sampling complicated posterior distributions, the efficiency of the Metropolis algorithm and the Gibbs algorithm can be highly deteriorated by randomness, especially in highly parameterized Bayesian inverse problems. Therefore, to collect sufficient samples for inference, the chain needs to run tens of thousands or even more iterations. Besides, as forward models of hydrogeological inverse problems are computationally expensive, the computational overhead of MCMC becomes unmanageable. The huge computational cost raises one of the most important concerns about the Metropolis algorithm, which is to improve its efficiency in exploring the posterior distribution, and to decrease the number of forward simulations during the inversion process. Such concerns prompt the development of new algorithms that incorporate the gradient of the loss function ℓ into the Metropolis to suppress the randomness of the exploration [*Cui et al.,* 2011; *Efendiev et al.,* 2005; *Fu and Gómez-Hernández,* 2009; *Martin et al.,* 2012].

One of the most favorable gradient-based MCMC methods is the Metropolis-adjusted Langevin algorithm (MALA), which are proposed by *Roberts and Tweedie* [1996]. Compared with the complete randomness within the Metropolis algorithm, MALA

proposes a sample by a combination of a deterministic step and a random-walk step. The deterministic step in MALA requires the gradient of the loss function with respect to the random variable, which can be regarded as a gradient descent step. After a gradient descent step is done, a scaled random walk step is implemented to the random variable, whose covariance matrix can be defined as the identity matrix or the prior covariance matrix. The new sample of the random variable is used as a proposal for the Markov chain. MALA can also push the performance further by taking advantage of the local manifold structure of the loss function [*Girolami and Calderhead,* 2011].

Another popular member in the MCMC society is the Hamiltonian Monte Carlo algorithm (HMC) [*Neal,* 2011]. HMC is also a member of Metropolis algorithms, which simulates the Hamiltonian dynamics that can be recognized as several gradient descent steps. The multi-step mechanism of HMC also contributes to an elimination of oscillation when the chain is running toward the posterior distribution. Thus, the convergence speed of HMC is usually faster than that of the MALA due to its higher efficiency. The efficiency of HMC can also be further improved by incorporating information of local manifold structures [*Girolami and Calderhead,* 2011], which is modified and applied by *Bui-Thanh and Girolami* [2014] to solve a large-scale Bayesian inverse problem. However, instead of iteratively computing the Gauss-Newton Hessian, which coincidently is the Fisher information matrix in this case, *Bui-Thanh and Girolami* [2014] applied a fixed Gauss-Newton Hessian of the maximum a posterior (MAP) point through the whole sampling process. The results indicate that the simplified RHMC is well capable of solving a large-scale Bayesian inverse problem.

Inspired by the idea of exploring local structures of the posterior, we also realized the importance of harnessing curvature information in a gradient-based MCMC algorithm. *Martin et al.* [2012] designed a Metropolis algorithm that proposes proposal samples imitating the Newton optimization algorithm, the stochastic Newton algorithm. To avoid the computational overhead of computing and storing the Hessian matrix, the authors also adopted a Hessian-free approach that is based on the Lanczos algorithm [*Lanczos,* 1950]. Owing to the efficient exploitation of the local curvature information, the stochastic Newton algorithm out performs the MALA algorithm in their numerical experiments.

In this thesis, by employing a PCA in the original variable space, the exact computation of a Gauss-Newton approximation of the Hessian matrix becomes practical, therefore, we combined the Gauss-Newton algorithm with the Metropolis algorithm to explore the posterior distribution.

2.3 Non-Ensemble-Based Bayesian Inverse Approach

Contrast to the ensemble-based Bayesian inverse approach, non-ensemble-based approach directly searches for the MAP estimate of the posterior distribution without simulating any conditional realizations, which is usually achieved by optimization algorithms. The associated uncertainty of the MAP estimate is mostly computed by its local covariance matrix. Therefore, non-ensemble-based approaches can hardly be employed in scenarios where further Monte Carlo simulations using conditional realizations are called for. However, the non-ensemble-based approach usually possess higher computational efficiency, as the number of forward model evaluations is usually lower using this approach.

Researchers have made continued dedications to improve the efficiency of these optimizations algorithms. *Hu et al.* [2011] proposes an approximated Hessian preconditioned conjugate gradient approach to solve seismic wave inversion problems. *Bui-Thanh et al.* [2012] proposes a conjugate gradient approach that combined with a randomized approximating algorithm for Hessian. This approach is also validated in a seismic wave Bayesian inverse problem.

In hydrogeology, by embracing a geostatistical description of the prior covariance matrix [*Matheron,* 1963], the geostatistical approach was proposed to solve Bayesian inverse problems. Two representative geostatistical approaches are the quasi-linear geostatistical approach and the Successive Linear Estimator (SLE). SLE is a cokriging-like geostatistical approach that produces a linear estimator with successive assimilation of nonlinear relationship in the forward model [*Yeh et al.,* 1995; *Yeh et al.,* 1996]. Its effectiveness and efficiency have been validated in various applications, especially in hydraulic tomography problems [*Liu et al.,* 2002; *Yeh and Liu,* 2000; *Zhu and Yeh,* 2005; *Zhu and Yeh,* 2006]. Recently, to expand the application realm of SLE, Karhunen-Loeve expansion is incorporated into it to construct a reduced-order SLE for large-scale hydraulic tomography problems [*Zha et al.,* 2018].

Among the members of non-ensemble-based approaches, this thesis mainly focuses on the geostatistical approach (GA) proposed by *Hoeksema and Kitanidis* [1984]; *Kitanidis and Vomvoris* [1983a]; *Kitanidis* [1995], which is an iterative quasi-linear approach. The quasi-linear geostatistical approach also has a close relation to Gauss-Newton optimization and cokriging [*Nowak and Cirpka,* 2004], and thus has high efficiencies in solving Bayesian inverse problems. However, for highly-parameterized Bayesian inverse problems

that admit high-resolution hydraulic conductivity fields, the geostatistical approach still calls for special implementations to abate the computation overhead [*Lee and Kitanidis,* 2014; *Liu and Kitanidis,* 2011]. The reasons lie in two aspects: firstly, numerical simulation of forward models needs to be performed on a high-resolution grid for multiple times; and secondly, the large number of unknowns increases the cost of matrix computation during the inversion.

In the past a few decades, many researchers have devoted to reducing the computational cost of the quasi-linear approach. *Nowak et al.* [2003] employed an FFT (the fast Fourier transform) algorithm for efficient computations of the cross-covariance matrix, which was later applied to a sandbox inverse problem [*Nowak and Cirpka,* 2006a]. *Liu and Kitanidis* [2011] proposed a sparse prior covariance matrix that can be recognized as a general smoothing operator, which facilitates the storage and the computation of the covariance matrix. This sparsely formulated approach was validated in a sandbox hydraulic tomography problem. For large-scale linear geostatistical inverse problems, *Ambikasaran et al.* [2013] and *Saibaba et al.* [2012] proposed hierarchical matrices to accelerate the solving process. For Bayesian inverse problems with large volume of observational data, *Klein et al.* [2017] proposed a preconditioned conjugate gradient algorithm, and *Lin et al.* [2017] proposed a randomized data reduction algorithm to condition the quasi-linear approach. For inversion of time series data, such as pumping and tracer tests, the concept of temporal moments was applied. Instead of running the transient forward model hundreds or even thousands of times, the temporal moment approach transforms them to be steady-state for faster forward model evaluations [*Li et al.,* 2005; *Nowak and Cirpka,* 2006a; *Pollock and Cirpka,* 2008; *Yin and Illman,* 2009; *Zhu and Yeh,* 2006].

Recently, *Kitanidis and Lee* [2014] introduced the principal component analysis (PCA) in the quasi-linear geostatistical approach to ease the computation of Bayesian inverse problems, named as principal component geostatistical approach (PCGA), and demonstrated its effectiveness and efficiency in several hydrogeological applications [*Kang et al.,* 2017; *Lee and Kitanidis,* 2014; *Lee et al.,* 2016]. PCGA substantially reduces the computational cost associated with the covariance matrix and the computation of the Jacobian matrices. Besides, unlike other optimization approaches that rely on adjoint-solvers, PCGA only requires forward model itself to compute the Jacobian matrix.

CHAPTER 3. STOCHASTIC GAUSS-NEWTON WITH BROYDEN METHOD

3.1 Introduction

Application of inverse modeling in hydrogeology is universal and its necessity is mostly justified by providing estimates of hydraulic conductivity, the key parameters and the major uncertainty source in groundwater flow and transport systems [*Carrera et al.*, 2005; *Carrera and Neuman*, 1986; *Kitanidis*, 1996; *Luo and Cirpka*, 2011; *McLaughlin and Townley*, 1996; *Vrugt et al.*, 2008; *Yeh*, 1986; *Zimmerman et al.*, 1998]. Subject to the fine domain discretization for computation of heterogeneous fields, the number of unknown hydraulic conductivities can easily grow up to thousands or even millions, in contrast to a limited number of hydraulic measurements. As a consequence, inverse problems of inferring hydraulic conductivities are often highly ill-posed and irritatingly sensitive to even minor changes in the measurements [*Chen et al.*, 2008].

Inverse problems can be handled by various computational methods. One universally supported way of formulating an inverse problem is achieved by a Bayesian framework, which relies on depicting the characteristics of the posterior distribution and is well suited for problems with uncertainties from multiple measuring sources, scales, and complex forward models[*Carrera and Neuman*, 1986; *Kitanidis and Vomvoris*, 1983b; *Liu and Kitanidis*, 2011; *Michalak and Kitanidis*, 2003; *Rubin et al.*, 2010; *Tsai and Li*, 2008].

As the Bayes' rule states, posterior distributions are proportional to the product of a likelihood function and a prior distribution, and therefore constrained by the information

from available data and prior knowledge. One way to characterize the posterior distribution is to provide maximum-a-posteriori (MAP) estimate [*Carrera and Neuman,* 1986; *Zhang et al.,* 2014], if the posterior has a well-defined peak. MAP estimate can be obtained with minor computing power for moderate number of unknowns due to its comparatively low computational demand. However, for cases with unknown structural parameters, e.g., the spatial correlation of state variables, MAP estimates including all prior unknown information may become biased [*Kitanidis,* 1996]. The geostatistical Bayesian method proposed by *Kitanidis* [1995, 1996]; *Kitanidis and Vomvoris* [1983b] estimates structural parameters separately and provides unbiased best estimates and associated variances for linear or quasi-linear problems [*Fienen et al.,* 2008; *Li and Cirpka,* 2006; *Li et al.,* 2007; *Li et al.,* 2008; *Liu and Kitanidis,* 2011; *Nowak and Cirpka,* 2006b].

However, a single point estimate might lose its validity under circumstances when the posterior distribution does not possess a single-mode pattern. Therefore, more sophisticated and delicate statistical tools are required to delineate its detailed features, among which Markov Chain Monte Carlo (MCMC) stands out. Adopting a different philosophy of approximating the posterior, MCMC depicting the characteristics of the posterior distribution by simulating conditional realizations from a Markov chain [*Bates and Campbell,* 2001; *Efendiev et al.,* 2005; *Fu and Gómez-Hernández,* 2009; *Vrugt and Ter Braak,* 2011; *Vrugt et al.,* 2009]. MCMC produces more reliable estimates for nonlinear Bayesian inverse problems as it slacks the assumption of a Gaussian posterior, but is more computationally intensive compared to the geostatistical Bayesian approach and other optimization-based approaches [*Tonkin and Doherty*, 2005, 2009].

In the past decades, numerous efforts have been dedicated to a more computationally affordable MCMC for high dimensional Bayesian inverse problems, which can be roughly categorized into three different focuses. The first one focused on reducing dimensions of unknown variables by compressing the prior distribution. As prior models of the unknown variables in hydrogeology are mostly defined using geostatistical models, either two-point based or multiple-point based, machine learning algorithms are well-suited to simplify them [*Aharon et al.,* 2006; *Schölkopf et al.,* 1997; *Wold et al.,* 1987]. Former researchers exploited the ideas of principal component analysis (PCA) [*Kitanidis and Lee,* 2014], sparse coding approach [*Khaninezhad and Jafarpour,* 2014], discrete cosine transformation [*Jafarpour and McLaughlin,* 2009] and kernel PCA [*Sarma et al.,* 2008] to accelerate their inverse computation. The second one falls in reducing the complexity or computational cost of forward models, with acceptable approximation errors. Mixed-scale forward models were employed to reduce computational costs by running forward models on coarser grid resolutions [*Cui et al.,* 2011; *Efendiev et al.,* 2005; *Mondal et al.,* 2010]. Sparse-grid based interpolation was implemented to construct simplified surrogate models [*Liao et al.,* 2017; *Zeng et al.,* 2012]. The third one focused on devising more effective exploration algorithms for the posterior distribution in large-scale Bayesian inverse problems. By exploiting local structures of the likelihood function, less correlated proposals with higher acceptance rate were constructed by Metropolis algorithm [*Cui et al.,* 2014; *Law,* 2014]. The relationship between optimization algorithms and MCMC was also investigated and utilized to improve sampling efficiency of MCMC [*Martin et al.,* 2012; *Vrugt and Ter Braak,* 2011; *Vrugt et al.,* 2009]. Adaptive approaches were proposed

to improve the efficiency of Hamiltonian Monte Carlo (HMC) [*Girolami and Calderhead,* 2011; *Hoffman and Gelman,* 2014].

In this chapter, we base a faster MCMC algorithm on the stochastic Newton algorithm for solving Bayesian inverse problems in hydrogeological applications. This approach incorporates the Broyden algorithm for a much cheaper evaluation of Jacobian matrix between iterations. Numerical experiments corroborates the superiority of this approach by detailed analysis of accuracy of inversion results, computational time.

3.2 Principal Component Analysis

In hydrogeological inverse problems, because of the dense discretization of the random field, the space of the interested random variable is usually high dimensional, and therefore makes the inversion process computationally infeasible. The main idea of model reduction is to construct a map, which can be explicitly defined or implicitly defined, from a high-dimensional model with a dense covariance matrix to a low-dimensional model with a simpler covariance matrix. Instead of being solved in the original space, the inverse problems can be solved in the reduced space, and transformed back into the original one. Thus, the map is supposed to provide a way of reconstructing the original variables via the reduced variables. This chapter accepts PCA as the model reduction tool, also known as Karhunen-Loeve expansion in other literatures, to simplify the prior models that admits a dense covariance matrix [*Elsheikh et al.,* 2014; *Kitanidis and Lee,* 2014; *Saley et al.,* 2016; *Tompkins et al.,* 2010].

PCA generates a group of orthogonal base vectors, on which the realizations vary the most. The significant vectors, which are also known as principal components, are obtained via the transformation of the covariance matrix.

$$\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{T}} = [\mathbf{v}_1, \mathbf{v}, \dots, \mathbf{v}_{\mathrm{m}}] \begin{bmatrix} \mathrm{D}_1 & 0 & 0 & 0 \\ 0 & \mathrm{D}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathrm{D}_{\mathrm{m}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^{\mathrm{T}} \\ \mathbf{v}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_{\mathrm{m}}^{\mathrm{T}} \end{bmatrix} \tag{12}$$

where $\mathbf{D} \in \mathbb{R}^{\mathrm{m}\times\mathrm{m}}$ is a diagonal matrix with diagonal element D_{i} representing the ith largest eigenvalue of the matrix $\mathbf{C}$, $\mathbf{v}_{\mathrm{i}} \in \mathbb{R}^{\mathrm{m}\times 1}$ in matrix $\mathbf{V} \in \mathbb{R}^{\mathrm{m}\times\mathrm{m}}$ is its corresponding eigenvector, which is also called as principal component in the context of PCA.

The principal component $\mathbf{v}_{\mathrm{i}}$ can be viewed as a linear combination of the stochastic part in $\mathbf{s}$ that accounts for $\frac{\mathrm{D}_{\mathrm{i}}}{\Sigma_{\mathrm{i}=1}^{\mathrm{m}}\mathrm{D}_{\mathrm{i}}}$ percent of the total variance. To reduce the dimension of the covariance matrix, the matrix $\mathbf{V}$ is truncated to keep only the first $\mathrm{k} \ll \mathrm{m}$ principal components, but still account for most of the total variance of $\mathbf{s}$. The covariance matrix can then be simplified as $\mathbf{C} \approx \mathbf{V}_{\mathrm{k}}\mathbf{D}_{\mathrm{k}}\mathbf{V}_{\mathrm{k}}^{\mathrm{T}}$ with minor loss in the total variance, where $\mathbf{D}_{\mathrm{k}} \in \mathbb{R}^{\mathrm{k}\times\mathrm{k}}$ represents a diagonal matrix of the first k largest eigenvalues. Using truncated principal components $\mathbf{V}_{\mathrm{k}}$, $\mathbf{s}$ can be approximated as:

$$\mathbf{s} = \boldsymbol{\mu}_{\mathbf{s}} + \mathbf{V}_{\mathrm{k}}\mathbf{a} \tag{13}$$

where $\mathbf{a} \in \mathbb{R}^{\mathrm{k}\times 1}$ represents principal component coefficients. Any direct inversion of $\mathbf{s}$ can be circumvented by only estimating $\mathbf{a}$. As k is a much smaller number compared with m,

the computational gain is significant. Assume $\mathbf{a}$ has a normal prior $\mathbf{a}\sim\mathrm{N}(\boldsymbol{\mu}_\mathbf{a}, \boldsymbol{\Sigma}_\mathbf{a})$. The mean and covariance matrix of $\mathbf{a}$ can be computed as

$$\boldsymbol{\mu}_\mathbf{a} = \mathrm{E}[\mathbf{a}] = \mathrm{E}\left[\mathbf{V}_\mathrm{k}^\mathrm{T}(\mathbf{s} - \boldsymbol{\mu}_\mathbf{s})\right] = \mathbf{0} \tag{14}$$

and

$$\begin{aligned}\mathbf{C} = \mathrm{E}[(\mathbf{s} - \boldsymbol{\mu}_\mathbf{s})^\mathrm{T}(\mathbf{s} - \boldsymbol{\mu}_\mathbf{s})] \approx \mathrm{E}[(\mathbf{V}_\mathrm{k}\mathbf{a})(\mathbf{V}_\mathrm{k}\mathbf{a})^\mathrm{T}] = \mathbf{V}_\mathrm{k}\mathrm{E}[\mathbf{a}\mathbf{a}^\mathrm{T}]\mathbf{V}_\mathrm{k}^\mathrm{T} \\ = \mathbf{V}_\mathrm{k}\boldsymbol{\Sigma}_\mathrm{a}\mathbf{V}_\mathrm{k}^\mathrm{T}\end{aligned} \tag{15}$$

According to $\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^\mathrm{T}$, it can be derived that

$$(\mathbf{V}^\mathrm{T}\mathbf{V}_\mathrm{k})\boldsymbol{\Sigma}_\mathbf{a}(\mathbf{V}^\mathrm{T}\mathbf{V}_\mathrm{k})^\mathrm{T} \approx \mathbf{D} \tag{16}$$

As all the principal components are orthogonal vectors

$$\begin{bmatrix}\boldsymbol{\Sigma}_\mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \approx \mathbf{D} = \begin{bmatrix}\mathbf{D}_\mathrm{k} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathrm{m}-\mathrm{k}}\end{bmatrix} \tag{17}$$

where $\mathbf{D}_{\mathrm{m}-\mathrm{k}} \in \mathbb{R}^{(\mathrm{m}-\mathrm{k})\times(\mathrm{m}-\mathrm{k})}$ represents a diagonal matrix of the rest $\mathrm{m}-\mathrm{k}$ eigenvalues. If $\boldsymbol{\Sigma}_\boldsymbol{a}$ is chosen to be $\mathbf{D}_\mathrm{k}$, the corresponding approximation error should be

$$\|\mathbf{D} - \boldsymbol{\Sigma}_\mathbf{a}\|_\mathrm{F} = \|\mathbf{D}_{\mathrm{m}-\mathrm{k}}\|_\mathrm{F} = \sqrt{\mathrm{D}_{\mathrm{k}+1}^2 + \mathrm{D}_{\mathrm{k}+2}^2 + \cdots + \mathrm{D}_\mathrm{m}^2} \tag{18}$$

where $\|\ \|_\mathrm{F}$ represents the Frobenius norm. The approximation error is the second norm of the rest $\mathrm{m}-\mathrm{k}$ eigenvalues, and because of the small magnitude of these eigenvalues, the error is almost trivial. By rescaling the eigenvectors $\mathbf{V}_\mathrm{k}$, the prior of $\mathbf{a}$ can be simplified to

be $\mathbf{N}(\mathbf{0},\mathbf{I})$. Using the reduced prior model and Bayes' rule, the posterior distribution of $\mathbf{a}$ can be formulated as:

$$\mathrm{p}(\mathbf{a}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\big(\mathbf{y}-\mathbf{f}(\mathbf{a})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y}-\mathbf{f}(\mathbf{a})\big)-\frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a}\right) \tag{19}$$

3.3 Stochastic Gauss-Newton

Gradient-based MCMC methods, which incorporates the local structure to suppress the random walk behavior of the Metropolis algorithm, are particularly effective for Bayesian inverse problems. Various gradient-based MCMC methods have been developed in the past a few decades, which include but are not limited to HMC [*Neal*, 2011], MALA [*Roberts and Tweedie*, 1996] and the Stochastic Newton (SN) [*Martin et al.*, 2012], etc.

Assume the corresponding lost function of the posterior distribution can be formulated as

$$\ell(\mathbf{a}) = \frac{1}{2}\big(\mathbf{y}-\mathbf{f}(\mathbf{a})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y}-\mathbf{f}(\mathbf{a})\big)+\frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a} \tag{20}$$

As stated in Chapter 2, finding the MAP estimator equals finding the minimum of the corresponding loss function. To generalize it, the simulated samples from the posterior are supposed to fall into a certain range around the minimum of the loss function. Therefore, an efficient sampler can be designed by exploiting the power of optimization algorithms. Given a point $\mathbf{a}_{\mathrm{i}}$ on the Markov chain, the loss function can be reformulated as

$$\ell(\mathbf{a}) = \frac{1}{2}(\mathbf{a} - \mathbf{a}_\mathrm{i} + \mathbf{A}^{-1}\mathbf{g})^\mathrm{T}\mathbf{A}(\mathbf{a} - \mathbf{a}_\mathrm{i} + \mathbf{A}^{-1}\mathbf{g}) + \text{const.} \tag{21}$$

where $\mathbf{g}$ represents the gradient $\frac{\partial \ell}{\partial \mathbf{a}} = -\mathbf{H}_\mathbf{a}^\mathrm{T}\mathbf{R}^{-1}\big(\mathbf{y} - \mathbf{f}(\mathbf{a})\big) + \mathbf{a}$, $\mathbf{A}$ represents the Hessian matrix $\frac{\partial \ell^2}{\partial^2 \boldsymbol{a}} = \mathbf{H}_\mathbf{a}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}_\mathbf{a} - \frac{\partial^2 \mathrm{f}}{\partial \mathbf{a}^2}\mathbf{R}^{-1}\big(\mathbf{y} - \mathbf{f}(\mathbf{a})\big) + \mathbf{I}$ and $\mathbf{H}_\mathbf{a}$ represents the Jacobian matrix $\frac{\partial \mathrm{f}}{\partial \mathbf{a}}$. As the posterior distribution is proportional to $\exp\big(-\ell(\mathbf{a})\big)$, the equation above indicates that $\mathbf{a}_\mathrm{i}$ is approximately from a Gaussian distribution with the mean of

$$\boldsymbol{\mu}_{\mathbf{a}_\mathrm{post}} = \mathbf{a}_\mathrm{i} - \mathbf{A}^{-1}\mathbf{g} \tag{22}$$

which, from an optimization perspective, is a Newton optimization step; with the covariance matrix of

$$\mathbf{C}_{\mathbf{a}_\mathrm{post}} = \mathbf{A}^{-1} \tag{23}$$

which is the inverse local Hessian matrix. The similarity in analytical forms motivates a natural way of finding a proposal sample

$$\mathbf{a} \sim \mathbf{N}(\mathbf{a}_\mathrm{i} - \mathbf{A}^{-1}\mathbf{g}, \mathbf{A}^{-1}) \tag{24}$$

where $\mathbf{a}_\mathrm{i}$ represents the current sampling point on the Markov chain. *Martin et al.* [2012] analyzed this connection between the Metropolis algorithm and the Newton optimization algorithm, and consequently proposed an efficient MCMC sampling approach based on a cheap approximation of the Hessian. However, instead of approximating the exact Hessian by abandoning its smallest eigenvalues, this thesis adopt another approximation of Hessian,

which is the Gauss-Newton approximation by neglecting the second-order term containing $\frac{\partial^2 \mathrm{f}}{\partial \mathbf{a}^2}$

$$\mathbf{A}_{\mathrm{GN}} = \frac{\partial \ell^2}{\partial^2 \boldsymbol{a}} = \mathbf{H}_{\mathbf{a}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}_{\mathbf{a}} + \mathbf{I} \tag{25}$$

In cases where $\mathbf{R} = \sigma^2 \mathbf{I}$, by adjusting the value of σ^2, it can also act like a Levenberg-Marquardt Hessian matrix. In some cases, the substitution of the exact Hessian with the Gauss-Newton Hessian approximation even outperforms the Newton optimization algorithm [*Chen,* 2011]. However, it should be notified that even this algorithm is mentioned as stochastic Gauss-Newton to reflect our use of the Gauss-Newton approximation of Hessian matrix, it is essentially the same algorithm as stochastic Newton. Owing to the reduced parameter space, the exact formulation of the Gauss-Newton Hessian matrix is computationally feasible. Besides, the computation of each column in the Jacobian matrix $\mathbf{H}_{\mathbf{a}}$ can be achieved with just two forward model runs

$$(\mathbf{H}_{\mathbf{a}})_{\mathrm{i}} = \mathbf{H}\mathbf{v}_{\mathbf{i}} = \frac{\mathbf{f}(\mathbf{s} + \mathbf{v}_{\mathrm{i}} * \delta) - \mathbf{f}(\mathbf{s})}{\delta} \tag{26}$$

where δ is small perturbation.

3.4 Quasi-Newton Method - Broyden Method

The above approach costs $\mathrm{k} + 1$ forward model runs in each sampling iteration. In a typical nonlinear large-scale hydraulic tomography problem, the computational overhead on running forward simulators are still considerably huge. To reduce the number of forward model runs further, a quasi-Newton, Broyden method, is embraced to compute the Jacobian

$\mathbf{H_a}$. Broyden method was initially designed to approximate the Jacobian matrix between two consecutive optimization steps [*Broyden,* 1965]. One significant advantage is that Broyden method requires just one forward model run for constructing the Jacobian matrix in each iteration.

Assume $\mathbf{a}_{i-1}$ is the latest state in the Markov chain, to move forward to a new state $\mathbf{a}_i$, the Jacobian matrix $\mathbf{H}_{\mathbf{a}_i}$ is in demand. Given the computational results from the ith iteration, which include the Jacobian matrix $\mathbf{H}_{\mathbf{a}_{i-1}}$, the forward model output $\mathbf{f}(\mathbf{a}_{i-1})$ and $\boldsymbol{\Delta}_i = \mathbf{a}_i - \mathbf{a}_{i-1}$, the new Jacobian $\mathbf{H}_{\mathbf{a}_i}$ can be computed via:

$$\mathbf{H}_{\mathbf{a}_i} = \mathbf{H}_{\mathbf{a}_{i-1}} + \frac{\left(\mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k\mathbf{a}_i) - \mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k\mathbf{a}_{i-1}) - \mathbf{H}_{\mathbf{a}_{i-1}}\mathbf{d}_i\right)\boldsymbol{\Delta}_i^{\mathrm{T}}}{\boldsymbol{\Delta}_i^{\mathrm{T}}\boldsymbol{\Delta}_i} \tag{16}$$

where $\mathbf{d}_i$ represents the deviation between $\mathbf{a}_{i-1}$ and $\mathbf{a}_i$. Broyden method can be recognized as a first-order approximation of the Jacobian matrix $\mathbf{H}_{\mathbf{a}_i}$ given $\mathbf{H}_{\mathbf{a}_{i-1}}$. According to this equation, by applying the Broyden method, the number of forward model runs in each optimization or sampling iteration is further reduced from $k + 1$ to one. That is, except for the computation of $\mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k\mathbf{a}_{i-1})$, all the other terms in Broyden method are from the former computation step. Therefore, by applying Broyden method to the sampler, the Jacobian matrices can be updated with no more forward runs compared to a simple random walk Metropolis algorithm. This combination is expected to benefit more for second-order gradient-based MCMC methods, as they are more computationally demanding.

It is worth noting that even though the Broyden's algorithm reduces the time complexity of computing the Jacobian matrix, it also raises the required storage as

$\mathbf{H}_{\mathbf{a}_i}$, $\mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k\mathbf{a}_{i-1})$ and $\boldsymbol{\Delta}_i$ need to be stored between iterations. For extremely huge datasets with n going up to $10^4 \sim 10^6$, the advantage Broyden method may be overcompensated by the large storage requirement.

3.5 Diagnostics of Convergence

At the beginning of MCMC, samples from the simulated Markov chains are not from the exact posterior distributions. Thus, a certain number of iterations are necessary for all the chains to converge to the targeted distribution before collecting samples from the chains. The iterations before convergence are called the 'burn-in' period. In this thesis, the multivariate potential scale reduction factor (MPSRF) computed based on samples from all the parallel chains is used to monitor and verify the convergence of the chains [*Gelman and Rubin,* 1992]. The threshold value of the MPSRF is prescribed to be 1.2. That is, we assume all the chains converge when the computed MPSRF is less than 1.2.

3.6 Numerical Experiments

In this section, we investigate the performance of the proposed sampling approach. The effectiveness of Broyden method is firstly evaluated to determine its sufficiency of approximating the Jacobian matrix. Then the proposed approach is compared with some other MCMC methods. The comparison is in terms of number of forward model runs, computational cost, and accuracy of inversion result.

Numerical experiments are designed to invert a 2-D heterogeneous hydraulic conductivity fields using multiple steady-state hydraulic head measurements. Synthetic natural log hydraulic conductivity fields are generated using an exponential covariance

function with structural parameters specified in Table 1. The true natural log field for this numerical experiment is shown in Figure 1. Values of steady-state hydraulic head are generated by numerically solving the governing equation of groundwater flow described in Table 1. Hydraulic head data used for inversion are collected from a monitoring well network represented by the white dots in Figure 1. The collected data are contaminated with 1%~2% normal random errors to serve as field measurements for Bayesian inversion. All numerical experiments are implemented on a desktop with Intel® Core™ i7-3770 CPU @ 3.40GHz and 32.GB RAM.

Table 1 Numerical experiment setup.

Geostatistical Properties	
Domain Scale	$l_x \times l_y = 10 \times 10$
Resolution	128×128
Covariance Model	Exponential model
Mean	$\mu = 0$
Variance	$\sigma^2 = 1$
Correlation length	$l_c = 5$
Governing Equation	
$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = 0$$	
Head Boundary Condition	
Top	Impermeable
Bottom	Impermeable
Left	$h\|_{x=0} = 0.1, 0.11, 0.12, \ldots, 0.15$
Right	$h\|_{x=l_x} = 0$

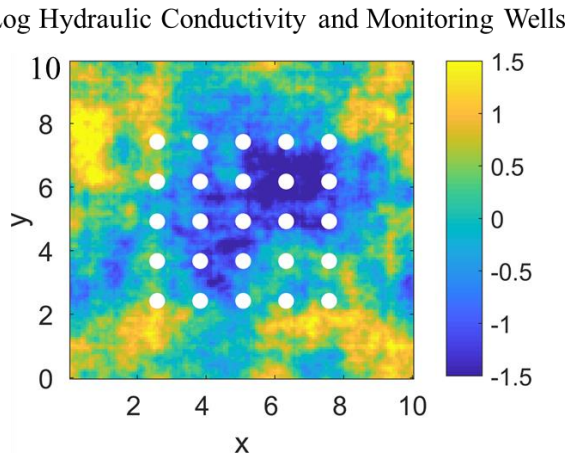

Figure 1 True field for the numerical experiment.

3.6.1 Effectiveness of Broyden Method in SGN

Broyden method is applied to the stochastic Gauss-Newton for a faster computation of gradient and approximated Hessian matrices during MCMC iterations, and it is called as "BSGN" as an abbreviation in the following for convenience. The performance of BSGN is compared with the standard SGN in the numerical experiment stated before. As PCA is adopted to reduce the number of random variables, a constant number of retained principal components are applied to both methods, which is $k = 30$.

Table 2 contrasts the computational cost of BSGN and the standard SGN. Both methods meet the convergence criteria, MPSRF less than 1.2, within 500 iterations, but the chains in both methods are advanced further for better mixing before they start to collect samples. The summary statistics indicates that Broyden method significantly reduces the computational cost in SGN, achieving a time reduction factor around 20, because Broyden method approximates the Hessian matrix and the gradient with only one forward model run in each MCMC iteration, while that number of the standard SGN is k.

Table 2 Statistics of computational performances of BSGN and SGN.

Properties	**BSGN**	**SGN**
Number of Iterations to Convergence	500	500
Number of Forward Runs to Convergence	3170	82246
Computational Time to Convergence (s)	727.22	13946
Total Number of Forward Runs	5808	164767
Total Computational Time (s)	1352.50	27886

Figure 2 compares the reproductions of the measurements for the two methods. The simulated head measurements are computed by 4 conditional realizations at the end of the Markov chains. Both methods produces tight fitting between simulated head values and measurements, which are numerically demonstrated by their low mean squared error (MSE).

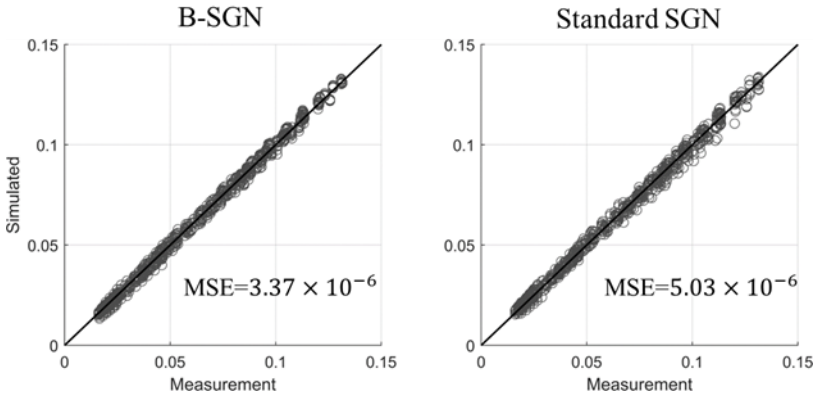

Figure 2 Reproduction of the measurements.

The inversion results using BSGN and SGN are shown in Figure 3, which include the best estimates based on ensemble means, and the histograms of the corresponding absolute errors. Both inversion results capture the dominant pattern of the true field. The quality of inversion results is also quantitatively analyzed by the mapping accuracy (MA), which represents the percentage of estimated points that have absolute error less than a threshold [*Kang et al.*, 2017]. In this section, the threshold is set to be 15% of the difference between the maximum and minimum of the true field, illustrated as the vertical red lines in Figure 3. That being given, the mapping accuracy can be evaluated by the percentage of points falling to the left of the threshold line. The mapping accuracy for BSGN and SGN are 83.37% and 81.96% respectively. The results inform that the Jacobian approximation by Broyden method does not jeopardize the quality of inversion result, and in this specific case, it even slightly improves the quality.

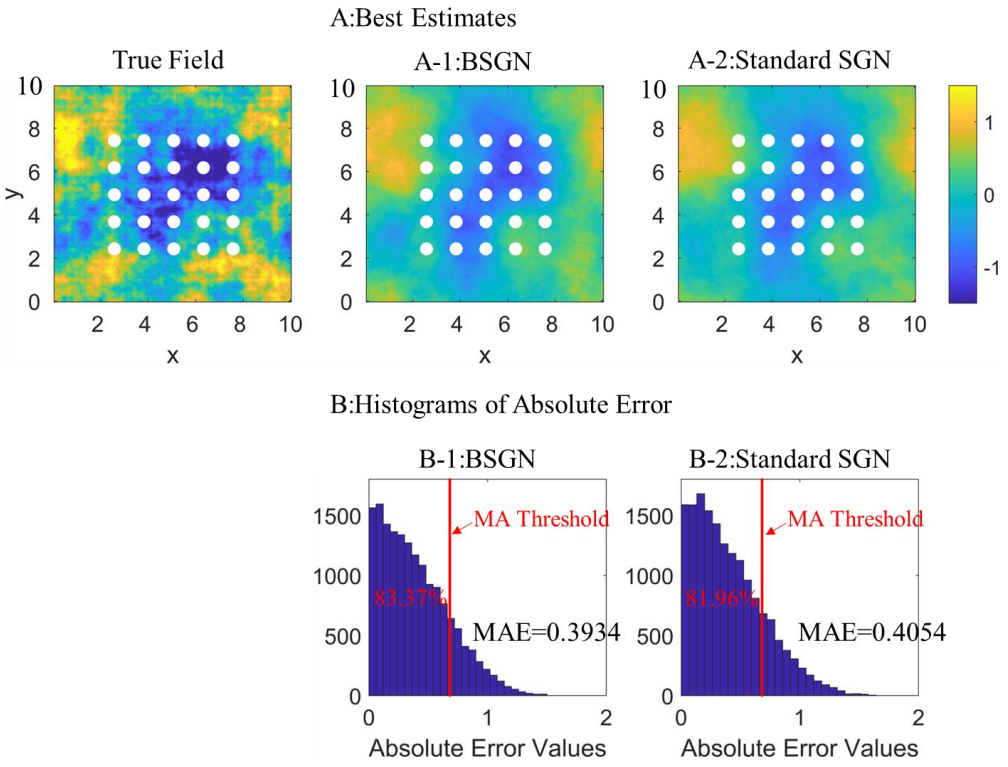

Figure 3 Inversion results using BSGN and SGN.

3.6.2 Compare BSGN with other MCMC algorithms

To compare the performance of BSGN with those of MALA and HMC in the PCA reduced space, the example in 3.6.1 is revisited. The same inverse problem is solved by BSGN, MALA and HMC. The number of retained principal components k is 30 for all algorithms. As MALA and HMC rely on an adjoint solver for efficient computation of gradients. An adjoint solver is developed accordingly, which requires almost the same computational cost as a forward solver. To set up MALA and HMC, the step size parameters in MALA and HMC need to be adjusted for the best sampling efficiency. In the MALA implementation, the suggestions in [*Roberts and Rosenthal*, 2001] is followed to adaptively adjust the step length every 100 iterations during the burn-in period and freeze it after the chain convergence. The achieved acceptance rate is around 54.7%~61.4%, which is very close to the optimal acceptance rate 57.4%. In the HMC implementation, a constant number of numerical simulation steps is adopted, which is 3 in this experiment, while the step length is adjusted every 100 iterations to achieve the optimal acceptance rate 65% as in [*Neal*, 2011]. Values of the step length are also frozen after the chain convergence. The achieved acceptance rate is around 65.9%~68.4%.

The computational summary is in Table 3. The statistics indicates that the proposed method needs much less iterations and number of forward model runs for convergence and finishing the entire sampling procedure. The total computational time of BSGN is less than one tenth of the HMC's and one twentieth of the MALA's.

Table 3 Summary of computational costs of BSGN, HMC and MALA.

Properties	BSGN	HMC	MALA
Number of Iterations to Convergence	500	3020	14140
Number of Forward Runs to Convergence	3170	84564	113124
Computational Time to Convergence (s)	727	14908	26127
Total Number of Forward Runs	5808	98564	117124
Total Computational Time (s)	1353	17399	26841
Number of Forward Runs per Iteration	5.8	28	8
Number of Forward Runs per Sample	2.9	49.3	58.6

The comparison of inversion results of the three methods are shown in Figure 4. All the three methods capture the main features of the true field. Considering the quantitative measure of the results, BSGN and HMC are on the same level, outperforming MALA for more than 5%. All methods provide tight fitting to the available measurements as shown in Figure 5, which proves their effectiveness in solving Bayesian inverse problems.

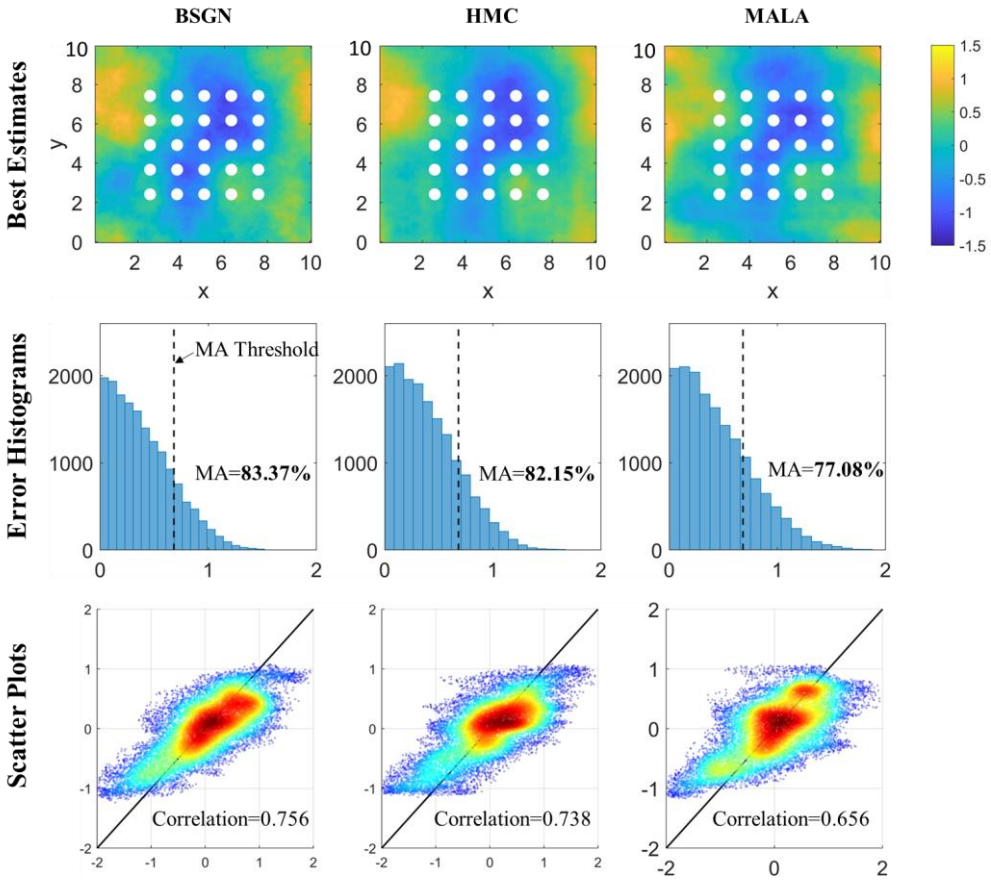

Figure 4 Comparison of inversion results between BSGN, HMC and MALA.

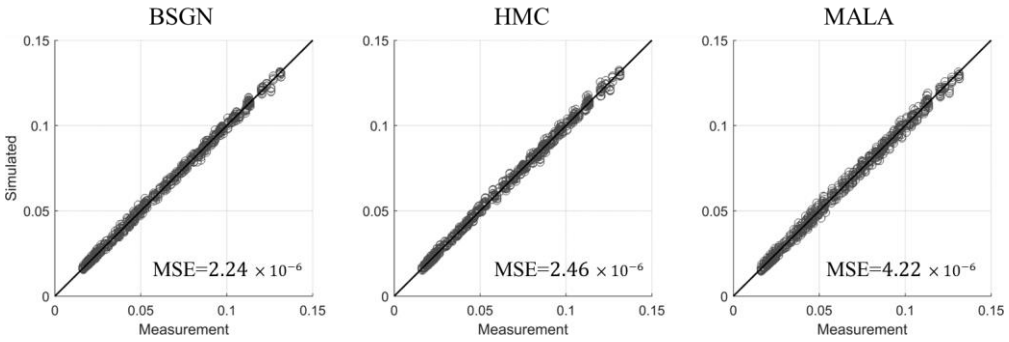

Figure 5 Reproduction of measurements.

To summarize, the proposed BSGN method outperforms MALA and HMC in a 2-D Bayesian inversion problem of steady state pumping data, which provides comparably high quality inversion results with faster convergence speed, lower computational cost, and higher sampling efficiency.

3.6.3 Computational Cost for Different Resolutions

The computational performance of BSGN is evaluated on problems with different resolutions. The physical model of these inverse problems is the same as the one stated at the beginning of this subsection. However, the random fields used for inversion also has different spatial patterns.

Figure 6 shows the computational time and the number of forward model runs for different problem resolutions. The circles represent the computational cost during the burn in period, while the dots represent the computational cost of the entire sampling process. Figure 6-a shows the relationship between the computational time and the problem resolution or the number of unknowns, m. Both x and y values in this figure are in log scale. It indicates that the computational time of BSGN is proportional to the inverse problem resolution for this specific case. Figure 6-b shows that the number of required forward model runs of BSGN, which does not depend on the problem resolution and stays around 6000 for all the tested cases, indicating that the overall computational time change primarily depends on the computational time change of each forward model run.

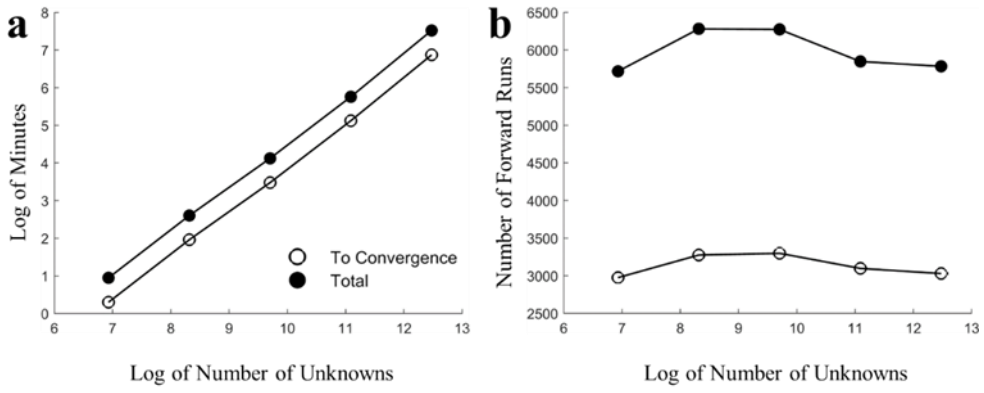

Figure 6 Computational time and numbers of forward model runs for different problem sizes using BSGN.

3.7 Conclusion

Broyden method, an algorithm designed for Jacobian approximation in optimization problems, is incorporated into SGN to improve its computational efficiency in solving hydrogeological Bayesian inverse problems. The newly developed method is investigated in a numerical experiment to demonstrate its effectiveness. The results indicate that the computational overhead is substantially cut down by Broyden method while a high accuracy of inversion results is preserved. This approach is also compared with other gradient-based MCMC algorithms in a PCA reduced variable space, which also strengthens the better performance of this approach.

We shall also notice that Broyden algorithm requires the storage of specific vectors and matrices after each iteration, which might become problematic for large-volume observational data. Consequently, special treatment or compression of the observational data are needed to lower the storage requirement.

CHAPTER 4. UPSCALING MCMC FOR HYDRAULIC TOMOGRAPHY

4.1 Introduction

Accurate characterization of spatially distributed hydraulic conductivities is crucial for reliable modeling of groundwater flow and solute transport in hydrogeology research. The local-scale characterization of hydraulic conductivities may be conducted by laboratory experiments based on core samples collected in the field [*Klute and Dinauer,* 1986a]. However, core samples can hardly be collected on an adequately discretized resolution to characterize a field site due to the high cost of well drilling. Therefore, inverse modeling is usually employed to infer the hydraulic conductivity field from aquifer testing measurements such as steady-state water head in monitoring wells for pumping tests, transient water head in monitoring and pumping wells for pumping or slug tests, and tracer concentration for tracer tests, etc. [*Cirpka et al.,* 2007; *Fienen et al.,* 2006; *Lee et al.*, 2016].

However, inverse modeling usually suffers from high ill-posedness that undermines the uniqueness and the stability of the solution. To meliorate the solution-finding process, constraints or regularizations are typically applied as priors in the Bayesian inverse framework. The strongest regularization, which also corresponds to the simplest inverse approach, is to assume a homogeneous hydraulic conductivity field, and analytical solutions are usually available as the forward model to evaluate the hydraulic heads or concentrations [*Kruseman et al.,* 1970]. Such a method takes a huge risk of oversimplification, and the estimate may only approximate the average property at the test

scale. Considering the complexity of hydrogeological formations, the conductivity field can be modeled as a spatially-distributed random function, described by either a training image or a two-point geostatistical covariance function, which encodes the spatial pattern of the hydraulic conductivity field [*Caers and Zhang*, 2004; *Liu*, 2006; *Minasny and McBratney*, 2005].

Aside from defining the prior information, large and informative datasets from aquifer tests always provide valuable information to improve the inversion accuracy and reduce the non-uniqueness. Hydraulic tomography, also known as sequential aquifer tests, has obtained growing attention in the past few decades [*Gottlieb and Dietrich*, 1995; *Liu et al.*, 2002].

Hydraulic tomography usually provides larger volume datasets than a traditional pumping test because a group of wells alternatively acts as a pumping well or monitoring wells to record the corresponding water heads. Moreover, due to the change of the pumping location, the collected head measurements are more informative as they are less correlated compared with a traditional single-well pumping test. Therefore, the tomographic layout also increases the information density of the dataset and reduces the non-uniqueness of the inverse solution. *Gottlieb and Dietrich* [1995] designed a numerical experiment to test the effectiveness of tomographic data in inferring permeability inhomogeneities. *Yeh and Liu* [2000] and *Zhu and Yeh* [2005] successfully inverted 3-D hydraulic conductivity fields using hydraulic tomographic data via a sequential approach. The effectiveness of hydraulic tomography was also verified in laboratory sandbox experiments [*Liu et al.*, 2002; *Liu and Kitanidis*, 2011].

Large-scale hydraulic tomography to estimate a high-resolution hydraulic conductivity field requires efficient inverse algorithms [*Lee and Kitanidis,* 2014; *Liu and Kitanidis,* 2011]. Firstly, to simulate the sequential aquifer tests, the forward model on a high-resolution field needs to be solved for multiple times. Secondly, the large number of unknowns increases the cost of matrix computation during the inversion and raises the computational overhead of each forward run. To reduce the number of unknowns, compression algorithms were recently introduced to inversion algorithms. *Kitanidis and Lee* [2014] used the principal component analysis (PCA) to reduce the dimension of the parameter space for Bayesian geostatistical inverse approach and demonstrated its application in a large-scale hydraulic tomography [*Lee and Kitanidis,* 2014]. Other dimension reduction algorithms are also available for hydrogeological inverse problems, such as kernel PCA [*Sarma et al.,* 2008], sparse coding approach [*Khaninezhad and Jafarpour,* 2014], discrete cosine transformation [*Jafarpour and McLaughlin,* 2009] or the state-of-art deep neural network [*Laloy et al.,* 2018]. In addition, to mitigate the computational expenses induced by the complexity of the forward model, researchers proposed surrogate model approaches, which construct a simpler surrogate model based on the input and output of the original one. Two of the many successful examples for groundwater applications are the Gaussian process approach [*Ju et al.,* 2018] and the sparse-grid based interpolation approach [*Liao et al.,* 2017; *Zeng et al.,* 2012]. By taking advantage of a delayed rejection Markov chain Monte Carlo (MCMC), coarse-scale or reduced-order forward models were adopted to reduce the computational cost associated with the large-scale forward model [*Cui et al.,* 2011; *Efendiev et al.,* 2005; *Mondal et al.,* 2010]. However, this approach only utilizes a coarse-grid or reduced-order forward model

in the acceptance-rejection stage. Thus, it merely saves computational resource on samples with low acceptance probability, but cannot fully avoid the computation of the large-scale forward model.

In this chapter, we propose an upscaling inverse approach to improve the efficiency of solving large-scale hydraulic tomography problems. Instead of running both the coarse-scale and fine-scale forward models, the proposed approach will fully circumvent the use of the fine-scale one. The original high-resolution field will be recovered by latent variables bridging the coarse-scale solution and the fine-scale solution, which are the principal component coefficients for this particular method. This chapter is organized as follows: PCA and the upscaling approach are firstly introduced, as they are the most essential part of our approach. Then the inversion algorithm, the stochastic Gauss-Newton (SGN) [*Martin et al.,* 2012] and the Broyden algorithm are introduced to make it a complete Bayesian inverse framework. A synthetic large-scale hydraulic tomographic dataset is constructed to test the effectiveness and the efficiency of the proposed approach. The performance is evaluated in terms of the reproduction of measurements, inversion accuracies and computational time.

4.2 Principal Component Analysis and Upscaling Approach

PCA has been introduced in the previous chapter as a model reduction approach, in this chapter, as the principal component coefficients are recognized as the latent variable, which connects the coarse-grid field and the fine-grid field for the upscaling approach, it will be briefly reviewed to compose a complete introduction of the upscaling approach.

A kernel trick is applied to implement the eigen-decomposition of the covariance matrix $\mathbf{C}$, which avoids a direct handling of this large and dense matrix. The decomposition results should be the same as expressed in Equation (12). The principal components are rescaled by their corresponding coefficients to transform $\mathbf{D}$ matrix into identity matrix $\mathbf{I}$, i.e., $\mathbf{v}_i = \mathbf{v}_i/D_i$, and there are only $k \ll m$ principal components retained in the matrix $\mathbf{V}$, the covariance matrix can be approximated as:

$$\mathbf{C} \approx \mathbf{V}_k \mathbf{I} \mathbf{V}_k^T = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix} \tag{27}$$

Therefore, the interested random variable $\mathbf{s}$ is able to be reconstructed using $\mathbf{V}_k$:

$$\mathbf{s} = \boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k \mathbf{a} \tag{28}$$

where $\mathbf{a} \in \mathbb{R}^{k \times 1}$ is a new random variable with prior distribution $\mathbf{N}(\mathbf{0}, \mathbf{I})$. Thus, the sampling process of finding appropriate $\mathbf{s}$ is substituted with finding samples of $\mathbf{a}$, which yield considerable computational advantage, as k is a much smaller number than m. The posterior distribution of $\mathbf{a}$ is given by:

$$p(\mathbf{a}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k \mathbf{a})\right)^T \mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{f}(\boldsymbol{\mu}_\mathbf{s} + \mathbf{V}_k \mathbf{a})\right) - \frac{1}{2} \mathbf{a}^T \mathbf{a}\right) \tag{29}$$

Even though the dimension of the interested variable space substantially shrinks after preprocessing of PCA, finding samples of the above distribution requires repeated

evaluations of $\mathbf{f}(\mathbf{s})$, which in most cases is an expensive numerical solver and may become problematic when the resolution of $\mathbf{s}$ goes up. However, by adopting the principal component coefficients as a group latent variable, as it is the variable to be inverted in the proposed approach, not the fine-grid field $\mathbf{s}$, each forward evaluation can be implemented on coarse-grid field, which can lower the computational cost of each forward run. Assume there exists an upscaling matrix $\mathbf{U}$ that satisfies the following relationship:

$$\mathbf{s}_2 = \mathbf{U}\mathbf{s} = \mathbf{U}\boldsymbol{\mu}_\mathbf{s} + \mathbf{U}\mathbf{V}_\mathrm{k}\mathbf{a} \tag{30}$$

where $\mathbf{s}_2 \in \mathbb{R}^{\mathrm{m}_2 \times 1}$ and m_2 is a much smaller number than m, the dimension of $\mathbf{s}$. Here after, the fine-resolution field is represented as $\mathbf{s}_1$ to distinguish the differences against the coarse-resolution field. To lower the computational cost, whenever a forward model run of $\mathbf{f}(\mathbf{s}_1)$ is requested, a forward model run of $\mathbf{f}(\mathbf{s}_2)$ could be used instead. The new forward relationship, or the measurement equation, is:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}_2) + \boldsymbol{\epsilon}_2 \tag{31}$$

where $\boldsymbol{\epsilon}_2$ represents errors raised by using $\mathbf{s}_2$ instead of $\mathbf{s}_1$ in forward model evaluations plus the original error term $\boldsymbol{\epsilon}$. The characteristics of $\boldsymbol{\epsilon}_2$ are supposed to be different from these of $\boldsymbol{\epsilon}$, as it is more of a systematic deviation. To compensate the influences caused by the approximation error, a theoretical sound way is to correct the Gaussian likelihood functions. However, according to our numerical experimental results in this chapter, the steady-state hydraulic head are sufficiently insensitive to the changes of simulation resolutions. That is, head simulation results are almost the same across different resolutions. Therefore, a simple strategy, the normalization of the available data $\mathbf{y}$, provides

high-quality inversion results in a traditional Gaussian likelihood framework. The posterior distribution is reformulated as

$$\mathrm{p}(\mathbf{a}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y}-\mathbf{f}(\mathbf{U}\boldsymbol{\mu}_{\mathbf{s}}+\mathbf{U}\mathbf{V}_{\mathrm{k}}\mathbf{a})\right)^{\mathrm{T}}\mathbf{R}'^{-1}\left(\mathbf{y}-\mathbf{f}(\mathbf{U}\boldsymbol{\mu}_{\mathbf{s}}+\mathbf{U}\mathbf{V}_{\mathrm{k}}\mathbf{a})\right)-\frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a}\right) \tag{32}$$

where $\mathbf{R}'$ is the covariance matrix of error term $\boldsymbol{\epsilon}_2$. In this note, $\mathbf{R}'$ is assumed to be equal to $\sigma^2\mathbf{I}$. To simulate samples of this distribution, only evaluations of $\mathbf{f}(\mathbf{s}_2)$ are demanded, which are distinctly cheap as they are solved on an upscaled field (the coarse-grid one).

As optimization-based MCMC algorithms are usually applied to solve the hydraulic tomography problem, the computation of the gradient vector and the Hessian matrix is described. Denote a loss function $\ell(\mathbf{a})$, the gradient is computed as $\mathbf{g}=\frac{\partial \ell}{\partial \mathbf{a}}$, while the Hessian is approximated as $\mathbf{A}=\frac{\partial^2 \ell}{\partial \mathbf{a}^2}$. Denote $\mathbf{U}\mathbf{V}_{\mathrm{k}}$ as a new matrix $\mathbf{V}'_{\mathrm{k}} \in \mathbb{R}^{\mathrm{m}_2\times \mathrm{k}}$, the gradient and the Hessian can be expressed as

$$\mathbf{g}_2=-\frac{1}{\sigma^2}(\mathbf{H}_2\mathbf{V}'_{\mathrm{k}})^{\mathrm{T}}\left(\mathbf{y}-\mathbf{f}(\mathbf{s}_2)\right)+\mathbf{a} \tag{33}$$

$$\mathbf{A}=\frac{1}{\sigma^2}(\mathbf{H}_2\mathbf{V}'_{\mathrm{k}})^{\mathrm{T}}(\mathbf{H}_2\mathbf{V}'_{\mathrm{k}})+\mathbf{I} \tag{34}$$

where $\mathbf{H}_2$ is the Jacobian matrix using the coarse-grid forward model, σ^2 is the variance in matrix $\mathbf{R}'$. The product of $\mathbf{H}_2\mathbf{V}'_{\mathrm{k}}$ is the Jacobian matrix of $\mathbf{f}(\mathbf{s}_2)$ with respect to $\mathbf{a}$, which

can be efficiently initialized in $k+1$ forward model runs and computed by Broyden method in the following iterations. The details will be explained in next subsection.

By adopting the upscaling approach, forward model runs using the fine resolution random field $\mathbf{s}$ are all bypassed. Additionally, in the inversion process, any matrix computation regarding matrix $\mathbf{V}_k$ is avoided. Instead, these matrix computations are achieved by matrix $\mathbf{V}_k'$, which is a much smaller matrix. Such changes are also expected to provide higher computational efficiency for Bayesian inversion.

4.3 Stochastic Newton and Broyden's Algorithm

In this chapter, the inversion of the hydraulic tomographic data is achieved by a MCMC algorithm, the Gauss-Newton modification of the stochastic Newton algorithm [*Martin et al.*, 2012]. The connection between MCMC and optimization algorithms is utilized to improve the efficiency of Metropolis moves in the variable space. Combined with the Gauss-Newton optimization approach, the proposal distribution for each Metropolis step is given by $\mathbf{a} \sim \mathbf{N}(\alpha(\mathbf{a}_i - \mathbf{A}^{-1}\mathbf{g}), \beta\mathbf{A}^{-1})$, where $\mathbf{a}_i$ represents the last accepted sample of the chain, $\mathbf{A}$ and $\mathbf{g}$ represent the local approximated Hessian and the gradient at $\mathbf{a}_i$, which can be computed using Equation (33) and Equation (34), and α and β are two tunable parameters as step sizes.

$\mathbf{H_a}$ represents the Jacobian matrix. $(\mathbf{H_a})_{ij} = \frac{\partial \mathbf{f}(\mathbf{s})_i}{\partial \mathbf{a}_j}$ is the first-order derivative of the forward model output $\mathbf{f}(\mathbf{s})$ with respect to $\mathbf{a}$ on the location where data $\mathbf{y}$ was collected. The computation of the Jacobian matrix on the first iteration can be achieved using the chain rule of partial derivatives:

$$\mathbf{H_a} = \frac{\partial \mathbf{f}}{\partial \mathbf{a}} = \frac{\partial \mathbf{f}}{\partial \mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{a}} = \mathbf{H}_2 \mathbf{V}'_{\mathrm{k}} = \mathbf{H}_2 [\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_{\mathrm{k}}]$$
$$= [\mathbf{H}_2 \mathbf{v}'_1, \mathbf{H}_2 \mathbf{v}'_2, \mathbf{H}_2 \mathbf{v}'_3, \dots \mathbf{H}_2 \mathbf{v}'_{\mathrm{k}}] \tag{35}$$

where $\mathbf{H}_2$ represents the Jacobian matrix with each element as $(\mathbf{H}_2)_{\mathrm{ij}} = \frac{\partial \mathbf{f}_{\mathrm{i}}}{\partial \mathbf{s}_{\mathrm{j}}}$, the first order derivative of the coarse-grid forward model output $\mathbf{f}$ with respect to $\mathbf{s}$ on the location where data $\mathbf{y}$ was collected. The ith column of $\mathbf{H_a}$ can be computed using two forward model runs:

$$(\mathbf{H_a})_{\mathrm{i}} = \mathbf{H}_2 \mathbf{v_i} = \frac{\mathbf{f}(\mathbf{s} + \mathbf{v}'_{\mathrm{i}} * \delta) - \mathbf{f}(\mathbf{s})}{\delta} \tag{36}$$

where δ is a very small number based on the computational precision. In this chapter, $\delta = 10^{-8}$ is adopted for all the implementations. The forward model run of $\mathbf{f}(\mathbf{s} + \mathbf{v}'_{\mathrm{i}} * \delta)$ needs to be evaluated once for each column in $\mathbf{H_a}$, which amounts to k forward model runs. Plus one forward model run of $\mathbf{f}(\mathbf{s}_2)$, the computation of $\mathbf{H_a}$ can be achieved with $\mathrm{k} + 1$ forward model runs.

The above approach costs $\mathrm{k} + 1$ forward model runs in each optimization or sampling iteration. In a typical nonlinear large-scale hydraulic tomography problem, the computational overhead on running forward simulators are still considerably huge. To reduce the number of forward model runs further, Broyden method is adopted to compute the Jacobian, which was initially designed to approximate the Jacobian matrix between two consecutive optimization steps [*Broyden*, 1965]. One significant advantage is that, unlike

the above method requiring $k+1$ forward model runs, Broyden method requires just one forward model run for constructing the Jacobian matrix.

Assume $\mathbf{a}_{i-1}$ is the latest state in the Markov chain, to move forward to a new state $\mathbf{a}_i$, the Jacobian matrix $\mathbf{H}_{\mathbf{a}_i}$ is in demand. Given the computational results from the ith iteration, which include the Jacobian matrix $\mathbf{H}_{\mathbf{a}_{i-1}}$, the forward model output $\mathbf{f}(\mathbf{a}_{i-1})$ and $\boldsymbol{\Delta}_i = \mathbf{a}_i - \mathbf{a}_{i-1}$, the new Jacobian $\mathbf{H}_{\mathbf{a}_i}$ can be computed via:

$$\mathbf{H}_{\mathbf{a}_i} = \mathbf{H}_{\mathbf{a}_{i-1}} + \frac{\left(\mathbf{f}(\boldsymbol{\mu}_s + \mathbf{V}_k\mathbf{a}_i) - \mathbf{f}(\boldsymbol{\mu}_s + \mathbf{V}_k\mathbf{a}_{i-1}) - \mathbf{H}_{\mathbf{a}_{i-1}}\mathbf{d}_i\right)\boldsymbol{\Delta}_i^{\mathrm{T}}}{\boldsymbol{\Delta}_i^{\mathrm{T}}\boldsymbol{\Delta}_i} \tag{37}$$

which is a first-order approximation of the Jacobian matrix $\mathbf{H}_{\mathbf{a}_i}$ given $\mathbf{H}_{\mathbf{a}_{i-1}}$. The only unknown parameter is the forward model output $\mathbf{f}(\boldsymbol{\mu}_s + \mathbf{V}_k\mathbf{a}_i)$. Thus, by applying the Broyden method, the number of forward model runs in each optimization or sampling iteration is further reduced from $k+1$ to one. As in the specific application of SGN, Broyden method updates the Jacobian with no more forward runs compared with a simple random walk Metropolis algorithm.

4.4 Application to Hydraulic Tomography

4.4.1 Problem Setup

The effectiveness and computational cost of the proposed upscaling inverse approach is investigated in a large-scale hydraulic tomography problem. We studied the numerical experiments in [*Kang et al.*, 2017; *Klein et al.*, 2017] and set up a numerical experiment with similar boundary conditions, geostatistical and hydraulic settings. A $\log_{10}$ hydraulic

conductivity random field is generated using a Gaussian covariance function whose structural parameters are listed in Table 4. The synthetic $\log_{10}$ hydraulic conductivity field is also shown in Figure 7. Black dots in the random field represent 25 different wells that will act as a pumping or monitoring well alternatively. The water head in the monitoring wells are recorded when the pumping well pumps to the steady state. The recorded heads are contaminated with normal random errors to serve as available measurements for inversion. The standard deviation of the normal random error is 1.5% of the true head.

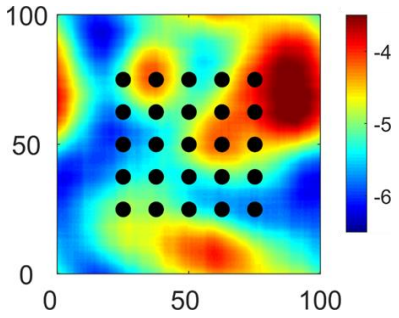

Figure 7 Parameters in forward model, log10 hydraulic conductivity and well setups.

Table 4 Geostatistical and hydrological parameters for the synthetic inverse problem.

Geostatistical Properties	
Domain Scale	$l_x \times l_y = 100\text{m} \times 100\text{m}$
Field Type	Log 10 Field
Resolution	512×512
Covariance Model	Gaussian model
Mean	$\mu = -5\text{m/s}$
Variance	$\sigma^2 = 0.5\text{m}^2/\text{s}^2$
Correlation length	$l = 20\text{m}$
Hydraulic Parameters	
Pumping Rate	$Q = 200\text{m}^3/\text{d}$
Governing Equation	
$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(\mathbf{K}\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = 0$	
Top	Impermeable
Bottom	Impermeable
Left	$h = 0.1\text{m}$
Right	$h = 0\text{m}$

4.4.2 Results and Discussion

The performance of the proposed inverse approach is investigated on the hydraulic tomography problem as proposed in the former subsection. This large-scale tomography problem is solved by the upscaling approach with forward solvers running on grid resolutions ranging from 128×128 to 16×16. In order to compensate the randomness involved in the inversion results, 10 different random fields are generated for the proposed hydraulic tomography problem and solved by the upscaling approach. The computational cost summary and quantitative analysis of the inversion results are based on the statistics of the 10 inversion results.

The total computational time and the number of forward runs are shown in Figure 8. Square data points represent the means values of the number of forward model runs, circle data points represent the mean values of the computational time. The error bars represent one standard deviation according to the statistics of the 10 inversion results. The error bars for the computational time are just too small to be visible. Along the upscaling direction, from high inversion resolution to low inversion resolution, the number of forward model runs does not change significantly, which means the number of forward evaluations of the upscaling approach does not depend on the target resolution. The rapid decrease in the total computational time along the upscaling direction is due to the lower computational time of the upscaled forward solvers. The highest reduction factor for the computational time can be approximated using the mean computational time, which is more than 2000 for this tomography problem. It demonstrates that the upscaling approach truly improves the computational efficiency for inversion.

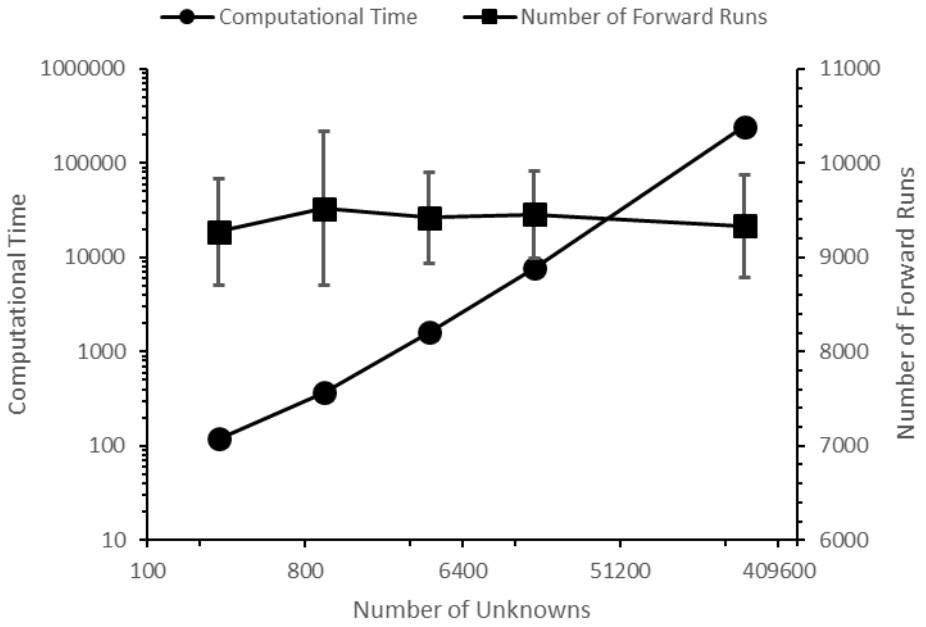

Figure 8 Computational time and numbers of forward model runs for the upscaled approach on different grid resolutions.

Besides the comparison of computational costs, the quality of inversion result needs to be investigated. The reproduction of the available measurements is firstly compared. The scatter plots of the comparison are shown in Figure 9. The simulated measurements are computed by the random fields generated at the end of the Markov chains. It can be found that by upscaling the forward model, the amount of points falling off the diagonal increases. The increasing misfit can also be reflected by the increasing MSEs along the direction of upscaling. The increasing misfit is mainly caused by the systematic deviation between the fine-grid forward model and the coarse-grid model.

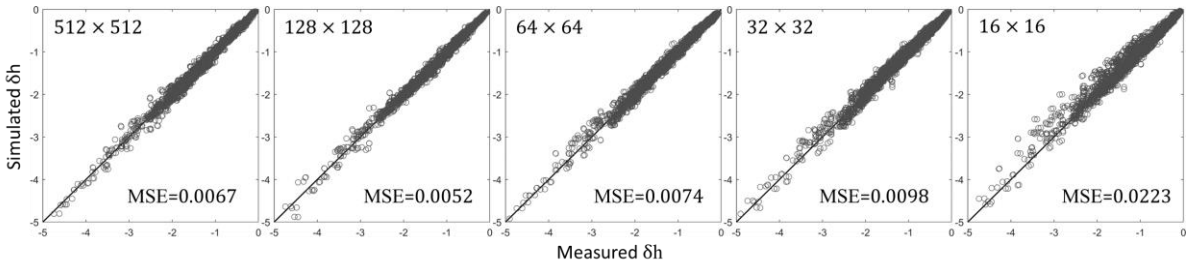

Figure 9 Reproduction of measurements.

The quality of the inversion result is further investigated by contrasting the estimated field with the original field. Inversion results of one random field, the best estimates of the $\log_{10}$ hydraulic conductivity and the histograms of their absolute error, are shown in Figure 10. The main features of the underlying true field are captured by the upscaling approach no matter which resolution it adopts for inversion. The total squared errors (TSEs) of the hydraulic conductivity field is used to quantify the accuracy of inversion. All the four upscaled inversion results provides low TSE values. Map accuracy (MA) is also adopted for accuracy quantification [*Kang et al.,* 2017]. MA stands for the percentage of the estimate that has absolute error less than a prescribed threshold [*Lee et al.,* 2016; *Yoon and McKenna,* 2012]. The threshold is set to be 15% of the difference between the maximum and the minimum of the true field. The threshold is shown as a vertical dash line in the absolute error histograms. By comparing absolute error histograms, it can be discovered that: 1) The inversion results of the upscaled approach provide similar and high MA values compared with the one solved on the original resolution. 2) In the case of 32×32 or 16×16 resolution, the proportion of the points with low absolute error decreases. The statement is confirmed with the smaller portions near zero in both histograms.

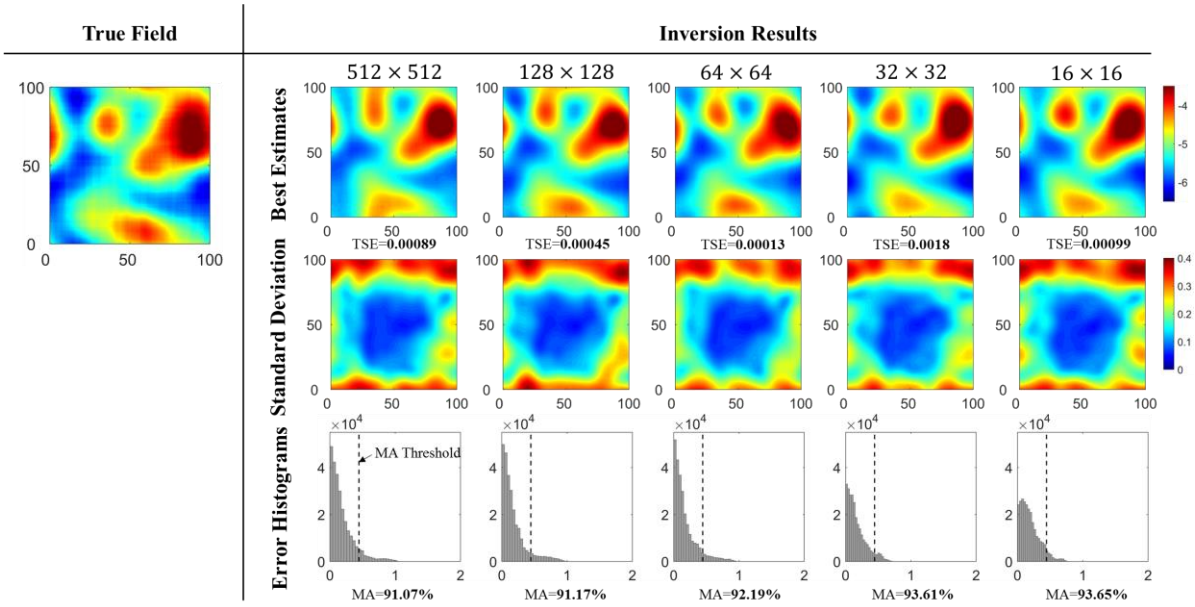

Figure 10 Comparison of inversion results of a single random field using forward solvers on different grids.

The statistics of TSE and MA values for the 10 different random fields is computed and plotted as curves in Figure 11. The dots represent the mean values of their TSEs and the squares represent the mean values of their MAs, while the error bars represent the values with one standard deviation away from the means. Along the direction of upscaling, the curves maintain high MA values and low TSE values, which indicate that the proposed upscaling approach solves the large-scale tomography problems with minor deterioration in the solution quality. Even if the forward solver is upscaled to the resolution of 16×16, the inversion results maintains low TSEs and high MAs. It also worth noting that the performance of the upscaling approach has larger standard deviations, which are reflected from the longer error bars in the figure.

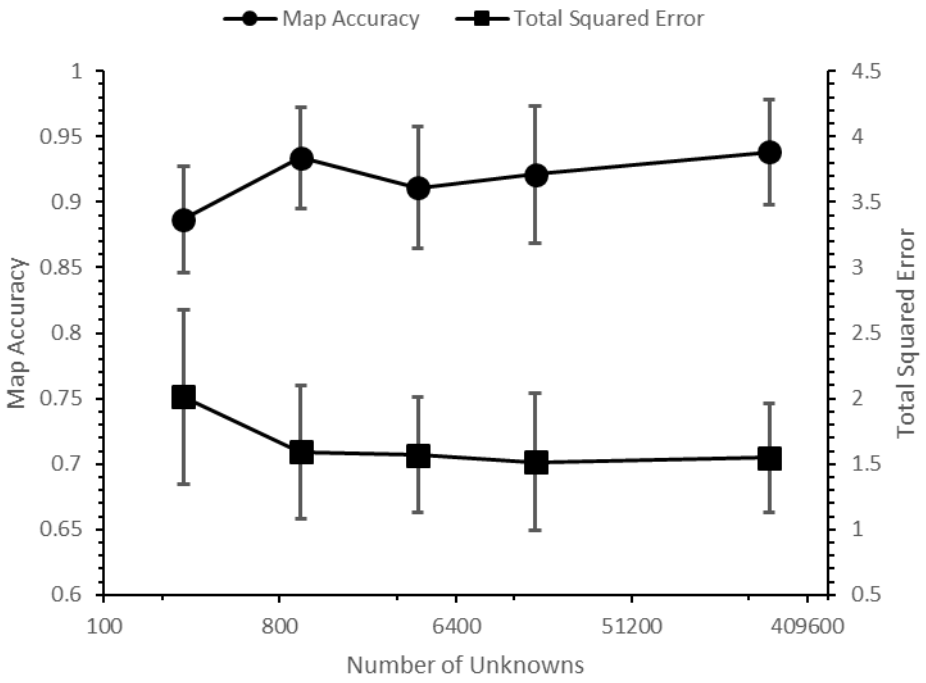

Figure 11 Inversion results of 10 different random fields.

To summarize this section, by running a numerical experiment with prescribed hydraulic and geostatistical parameters, we proved the efficiency, which is reflected by the low computational time, and the effectiveness, which is reflected by the high map accuracies and the low total squared error, of the upscaling approach in solving large-scale hydraulic tomography problems.

4.5 Conclusion

In this chapter, we proposed an upscaling approach for large-scale hydraulic tomography problems to overcome the computational overhead caused by expensive forward models. The essential idea is to construct a group of latent variables that can bridge the high-resolution and low-resolution variable. Therefore, the forward model can be solved with minor computational effort on the upscaled field. In our implementation, principal component coefficients are adopted as the latent variable. Besides, to reduce the number of forward model calls, Broyden's algorithm is adopted as it updates the Jacobian matrix with just one forward model call. The upscaling approach is incorporated into a stochastic Newton to make it a complete framework for Bayesian inversion.

The proposed approach is tested on a synthetic large-scale hydraulic tomography problem. According to the curves of the number of forward model runs, the upscaling inverse approach does not increase number of forward model runs for obtaining the inversion result. Because of using the upscaled forward solver, it significantly reduces the computational cost of solving the hydraulic tomography problem. The quality of the inversion results is assessed in terms of the reproduction of the available measurements, total squared error (TSE) and map accuracy (MA). The statistics of the results from ten

different random fields indicates that the proposed approach maintains almost the same level of inversion quality even on the coarsest resolution. The high efficiency and the inversion quality indicate that the upscaling approach has the potential to become a powerful tool for solving large-scale tomography problems.

One thing need to note is that the proposed approach has the potential to be generalized to other optimization based inverse approaches if an appropriate group of latent variable, which does not have to be principal component coefficients, can be found. However, the systematic error in the forward model output can be handled with a more statistically meaningful approach, which needs to be improved in the future work. There is one approach to be considered is an error-free inversion approach [*Rubin et al.*, 2010]. Besides, the upscaling approach may not be possible to be extended to channelized indicator fields, as the upscaling process will eliminate the sharp conductivity changes between different materials.

CHAPTER 5. REFORMULATION OF GEOSTATISTICAL APPROACH ON PRINCIPAL COMPONENTS

5.1 Introduction

Geostatistical Approach (GA) is an objective method that can provide rigorous inverse estimation of the best linear estimates and the associated uncertainties of unknown spatially-distributed parameter fields, for example, the heterogeneous hydraulic conductivity field, given indirect measurements such as hydraulic heads or concentrations. *Kitanidis* [1986] showed that traditional cokriging is equivalent to the estimation of the conditional mean of Gaussian processes in a Bayesian analysis. *Kitanidis* [1995] generalized the geostatistical approach in a rigorous Bayesian framework for solving quasilinear inverse problems to estimate spatially distributed parameters. A major problem of the quasilinear geostatiscal approach and cokriging techniques is that the computational costs are prohibitive for large-dimensional inverse problems. Major computational costs lie in handling auto-covariance matrices and cross-covariance matrices, including the storage and multiplication of matrices, and iterative implementation of forward models to determine the Jacobian matrix for nonlinear problems. Many efforts have been devoted to reduce the computational costs of the geostatistical approach. *Nowak et al.* [2003] proposed an efficient approach of computing the covariance matrix based on circulant embedding and the fast Fourier transform (FFT). *Liu and Kitanidis* [2011] proposed a sparse representation of the prior covariance matrix, which facilitates the storage and the associated computation. *Ambikasaran et al.* [2013] and *Saibaba et al.* [2012] proposed hierarchical matrices for large-scale linear geostatistical inverse problems. For

geostatistical inverse problems with massive observational data, *Klein et al.* [2017] proposed a preconditioned conjugate gradient algorithm, and *Lin et al.* [2017] proposed a randomized data reduction algorithm. Recently, *Kitanidis and Lee* [2014] introduced the principal component analysis (PCA) into the quasi-linear geostatistical approach and named the approach as principal component geostatistical approach (PCGA). PCGA employs a low-rank approximation of the covariance matrix, and by a so-called 'matrix-free' approach, the number of forward model runs for an explicit construction of the Jacobian matrix is reduced to almost the number of truncated principal components.

We shall notice that so far all the efforts have been devoted to advance the computational implementation of the geostatistical approach so that large-dimensional inverse problems can be solved within the classical framework of the quasi-linear geostatistical approach or cokriging. In this chapter, this issue is approached from a different perspective. We reformulate the inverse problem based on the principal component analysis of the unknown parameter field, which yields new unknown parameters of principal component coefficients instead of the original parameter field. This new approach is named as reformulated geostatistical approach (RGA). We demonstrate that the computational techniques introduced in PCGA by *Kitanidis and Lee* [2014] are assimilated naturally and intuitively in the new framework, and RGA unifies the problem setup and computational techniques. Moreover, compared with the textbook cokriging equations, the dimension of which depends on the number of observations, the number of normal equations to be solved for RGA is reduced to the number of truncated principal components, which also provides RGA more scalability, especially for inverse problems with massive observations.

We first reformulate the inverse problem in terms of principal component analysis of the parameter field. We then derive RGA solution for linear and quasilinear models and relate it to the classical geostatistical approach solutions. Two numerical experiments of hydraulic tomography are presented to demonstrate the validity of the reformulated approach.

5.2 Method

5.2.1 Reformulation of Inverse Problem

The general observation equation describing the relation between the data and unknowns is:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}) + \boldsymbol{\epsilon} \tag{38}$$

where $\mathbf{y} \in \mathbb{R}^{n\times 1}$ represents the observation data vector, $\mathbf{s} \in \mathbb{R}^{m\times 1}$ is the unknown variable vector, $\mathbf{f}$ represents the forward model, and $\boldsymbol{\epsilon} \in \mathbb{R}^{n\times 1}$ is Gaussian with mean $\mathbf{0}$ and covariance $\mathbf{R} \in \mathbb{R}^{n\times n}$ (typically proportional to the identity matrix). For the prior information, the random field $\mathbf{s}$ is Gaussian with an unknown mean and a generalized covariance function:

$$\begin{aligned} E[\mathbf{s}] &= \mathbf{X}\boldsymbol{\beta} \\ E[(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}] &= \mathbf{C} \end{aligned} \tag{39}$$

where $\mathbf{X} \in \mathbb{R}^{m\times p}$ represents the drift of mean, $\boldsymbol{\beta} \in \mathbb{R}^{p\times 1}$ represent the unknown coefficient vector of the drift function, and $\mathbf{C} \in \mathbb{R}^{m\times m}$ is the covariance matrix, typically evaluated by

a two-point geostatistical function in hydrologic and geophysical applications. p is typically a small number: for a constant mean, $p = 1$; and for a linear drift, $p = 2$.

The low-rank approximation of the symmetric covariance matrix, $\mathbf{C}$, can be written as the truncated eigen-decomposition (Lee and Kitanidis, 2014):

$$\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^T \approx \mathbf{V}_k\mathbf{V}_k^{\mathrm{T}} \tag{40}$$

The summation of all eigenvalues in the diagonal matrix $\mathbf{D}$ describes the total variance at all principal components. Thus, to determine the truncation number k, one may simply choose a k so that the ratio of the selected variance to the total variance is greater than a predefined criterion, such as 0.95, representing that 95 percent of variance can be described by the truncated principal components. Based on truncated principal components, $\mathbf{s}$ can then be approximated as:

$$\mathbf{s} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{V}_{\mathrm{k}}\mathbf{a} \tag{41}$$

where $\mathbf{a} \in \mathbb{R}^{k\times 1}$ is the unknown principal component coefficient vector, essentially quantifying the fluctuations of $\mathbf{s}$ with respect to the mean as a linear combination of principal components. Thus, any inversion of the stochastic part of $\mathbf{s}$ can be reduced to the inversion of $\mathbf{a}$. As k is independent of the number of observations, the inversion of $\mathbf{a}$ is more efficient for cases with massive observations. Moreover, the prior probability distribution of $\mathbf{a}$ is an i.i.d. standard multi-Gaussian distribution, $\mathbf{a}\sim\mathrm{N}(\mathbf{0},\mathbf{I})$. In fact, the above equation also provides a method to generate unconditional random fields of $\mathbf{s}$ given

mean and covariance. One can easily verify that the covariance of $\mathbf{s}$ generated is the low-rank approximation of $\mathbf{C}$.

The observation function can be reformulated as:

$$\mathbf{y} = \mathbf{f}(\mathbf{a}, \boldsymbol{\beta}) + \boldsymbol{\epsilon} \tag{42}$$

The posterior distribution of $\mathbf{a}$ and $\boldsymbol{\beta}$ can be formulated through Bayes theorem as:

$$p(\mathbf{a}, \boldsymbol{\beta}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\big(\mathbf{y} - \mathbf{f}(\mathbf{a}, \boldsymbol{\beta})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y} - \mathbf{f}(\mathbf{a}, \boldsymbol{\beta})\big) - \frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a}\right) \tag{43}$$

The best estimates of $\mathbf{a}$ and $\boldsymbol{\beta}$ are the maximum a posteriori (MAP) or the posterior mean values, which can be obtained by minimizing the negative logarithm of the posterior distribution:

$$\ell(\mathbf{a}, \boldsymbol{\beta}) = -\ln p(\mathbf{a}, \boldsymbol{\beta}|\mathbf{y}) \propto \frac{1}{2}\big(\mathbf{y} - \mathbf{f}(\mathbf{a}, \boldsymbol{\beta})\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{y} - \mathbf{f}(\mathbf{a}, \boldsymbol{\beta})\big) + \frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a} \tag{44}$$

5.2.2 Linear Model

For a linear relation between the observation and the unknown, we have:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \boldsymbol{\epsilon} = \mathbf{H}\mathbf{X}\boldsymbol{\beta} + \mathbf{H}\mathbf{V}_{\mathrm{k}}\mathbf{a} + \boldsymbol{\epsilon} \tag{45}$$

where $\mathbf{H} \in \mathbb{R}^{n\times m}$ is the sensitivity or Jacobian matrix. The function to be minimized becomes:

$$\ell(\mathbf{a},\boldsymbol{\beta}) \propto \frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{X}\boldsymbol{\beta}-\mathbf{H}\mathbf{V}_k\mathbf{a})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{X}\boldsymbol{\beta}-\mathbf{H}\mathbf{V}_{\mathrm{k}}\mathbf{a})+\frac{1}{2}\mathbf{a}^{\mathrm{T}}\mathbf{a} \tag{46}$$

By setting the derivatives with respect to $\mathbf{a}$ and $\boldsymbol{\beta}$ to zero, we can obtain two normal equations:

$$\begin{gathered}\hat{\mathbf{a}}^{\mathrm{T}}-\left(\mathbf{y}-\mathbf{H}\mathbf{X}\widehat{\boldsymbol{\beta}}-\mathbf{H}\mathbf{V}_{\mathrm{k}}\hat{\mathbf{a}}\right)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}=\mathbf{0}\\ \left(\mathbf{y}-\mathbf{H}\mathbf{X}\widehat{\boldsymbol{\beta}}-\mathbf{H}\mathbf{V}_{\mathrm{k}}\hat{\mathbf{a}}\right)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}=\mathbf{0}\end{gathered} \tag{47}$$

where $\hat{\mathbf{a}}$ and $\widehat{\boldsymbol{\beta}}$ are the best estimates of $\mathbf{a}$ and $\boldsymbol{\beta}$, respectively. Combining and rearranging the expressions leads to:

$$\begin{bmatrix}(\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{V}_{\mathrm{k}})+\mathbf{I} & (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X})\\ (\mathbf{H}\mathbf{X})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{Z}) & (\mathbf{H}\mathbf{X})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X})\end{bmatrix}\begin{bmatrix}\hat{\mathbf{a}}\\ \widehat{\boldsymbol{\beta}}\end{bmatrix}=\begin{bmatrix}(\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}\\ (\mathbf{H}\mathbf{X})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \tag{48}$$

This is a $k+p$ linear equation system, independent of the number of observations. Considering that the typical value of selected k is less than 100, the above equation is more efficient for large-dimensional problems with modest or large size of observations. The $(k+p)\times(k+p)$ matrix on the left side is the Hessian matrix of the objective function. Substituting $\hat{\mathbf{a}}$ and $\widehat{\boldsymbol{\beta}}$ into Equation (41) yields the best estimates of $\mathbf{s}$.

For $\mathbf{R}$ proportional to the identity matrix, $\mathbf{R}^{-1}$ in Equation (48) can be eliminated from both sides:

$$\begin{bmatrix}(\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}(\mathbf{H}\mathbf{V}_{\mathrm{k}})+\mathbf{R}_k & (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}(\mathbf{H}\mathbf{X})\\ (\mathbf{H}\mathbf{X})^{\mathrm{T}}(\mathbf{H}\mathbf{V}_{\mathrm{k}}) & (\mathbf{H}\mathbf{X})^{\mathrm{T}}(\mathbf{H}\mathbf{X})\end{bmatrix}\begin{bmatrix}\hat{\mathbf{a}}\\ \widehat{\boldsymbol{\beta}}\end{bmatrix}=\begin{bmatrix}(\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\mathbf{y}\\ (\mathbf{H}\mathbf{X})^{\mathrm{T}}\mathbf{y}\end{bmatrix} \tag{49}$$

where $\mathbf{R}_k \in \mathbb{R}^{k\times k}$ is the truncated error matrix proportional to the identity matrix.

The uncertainty of the estimate can be quantified the posterior covariance matrix of **a** and **β**, which can be computed by inverting the Hessian in Equation (48) and thus,

$$\begin{bmatrix} (\mathbf{HV}_\mathrm{k})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HV}_\mathrm{k}) + \mathbf{I} & (\mathbf{HV}_\mathrm{k})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HX}) \\ (\mathbf{HX})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HV}_\mathrm{k}) & (\mathbf{HX})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HX}) \end{bmatrix}^{-1} \tag{50}$$

This expression is generally computation-affordable since the rank of the matrix is $k + p$. The posterior covariance, jointly of **s** and **β**, can then be given by:

$$\begin{bmatrix} \mathbf{V}_\mathrm{k}{}^\mathrm{T} \\ \mathbf{X}^\mathrm{T} \end{bmatrix}^\mathrm{T} \begin{bmatrix} (\mathbf{HV}_\mathrm{k})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HV}_\mathrm{k}) + \mathbf{I} & (\mathbf{HV}_\mathrm{k})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HX}) \\ (\mathbf{HX})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HV}_\mathrm{k}) & (\mathbf{HX})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{HX}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{V}_\mathrm{k}{}^\mathrm{T} \\ \mathbf{X}^\mathrm{T} \end{bmatrix} \tag{51}$$

There is no need to evaluate the full matrix because we are interested in the diagonal terms, i.e., variances of the best estimates. We can complete the multiplication of the first two matrices, which yields an $m \times (k + p)$ matrix, and then only calculate the diagonal terms using corresponding row and column vectors [*Lee and Kitanidis*, 2014].

5.2.3 Quasilinear Model

For quasilinear inverse problems, we start with an initial guess of $\bar{\mathbf{a}}$ and improve the estimates at each iteration. The Jacobian matrix of **f** about **a** at $\bar{\mathbf{a}}$ is given by:

$$\bar{\mathbf{H}}_\mathbf{a} = \frac{\partial \mathbf{f}}{\partial \mathbf{a}}\bigg|_{\mathbf{a}=\bar{\mathbf{a}}} = \frac{\partial \mathbf{f}}{\partial \mathbf{s}}\bigg|_{\mathbf{s}=\bar{\mathbf{s}}} \frac{\partial \mathbf{s}}{\partial \mathbf{a}}\bigg|_{\mathbf{a}=\bar{\mathbf{a}}} = \bar{\mathbf{H}}\mathbf{V}_\mathrm{k} \tag{52}$$

where $\bar{\mathbf{H}}_{\mathbf{a}} \in \mathbb{R}^{n\times k}$ can be naturally determined by implementing the forward model k times to vary each component in $\mathbf{a}$. We shall notice that the 'matrix-free' approach introduced by *Kitanidis and Lee* [2014] is not needed or we can say such an approach has already becomes a 'built-in' feature because there is no need to determine $\bar{\mathbf{H}}$, i.e., the Jacobian matrix of $\mathbf{h}$ about $\mathbf{s}$ at $\bar{\mathbf{s}}$, and the matrix factorization is naturally built into $\bar{\mathbf{H}}_{\mathbf{a}}$.

Since $\mathbf{a}$ and $\boldsymbol{\beta}$ are separated, we also determine:

$$\bar{\mathbf{H}}_{\boldsymbol{\beta}} = \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{s}}\right|_{\mathbf{s}=\bar{\mathbf{s}}} \left.\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} = \bar{\mathbf{H}}\mathbf{X} \tag{53}$$

where $\bar{\mathbf{H}}_{\boldsymbol{\beta}} \in \mathbb{R}^{n\times p}$ requires to implement the forward model p times. By assuming that the actual $\hat{\mathbf{a}}$ and $\hat{\boldsymbol{\beta}}$ are close to $\bar{\mathbf{a}}$ and $\bar{\boldsymbol{\beta}}$, we approximate:

$$\mathbf{f}\left(\hat{\mathbf{a}}, \hat{\boldsymbol{\beta}}\right) = \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}(\hat{\mathbf{a}} - \bar{\mathbf{a}}) + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}\right) \tag{54}$$

Thus, the observation function can be reorganized as:

$$\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}} = \bar{\mathbf{H}}_{\mathbf{a}}\hat{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\,\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon} \tag{55}$$

Following the linear model presented, for each iteration, we can directly solve:

$$\begin{bmatrix} \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}} + \mathbf{I} & \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}} & \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \end{bmatrix} \tag{56}$$

The calculated $\hat{\mathbf{a}}$ is then substituted into equation (41) to evaluate $\hat{\mathbf{s}}$ for the forward model simulation for the next iteration. After the iteration procedure has converged, the best of estimate of $\mathbf{s}$ can be obtained, and the posterior covariance of $\mathbf{s}$ can be computed:

$$\begin{bmatrix}\mathbf{V_k}^{\mathrm{T}}\\ \mathbf{X}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}}+\mathbf{I} & \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}}\\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}} & \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{V_k}^{\mathrm{T}}\\ \mathbf{X}^{\mathrm{T}}\end{bmatrix} \tag{57}$$

5.3 Generating Conditional Realizations

RGA can also benefit the generation of conditional realizations, which are from a Gaussian approximation of the posterior distribution. After convergence, a Gaussian approximation of the posterior can be formulated by taking MAP as the mean and the inverse Hessian at MAP as the covariance matrix.

$$\begin{bmatrix}\mathbf{a}\\ \boldsymbol{\beta}\end{bmatrix}\sim\mathbf{N}\left(\begin{bmatrix}\hat{\mathbf{a}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix},\mathbf{C}_{\mathrm{posterior}}\right) \tag{58}$$

where $\mathbf{C}_{\mathrm{posterior}}$ can be computed as

$$\mathbf{C}_{\mathrm{posterior}}=\begin{bmatrix}\bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}}+\mathbf{I} & \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}}\\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\mathbf{a}} & \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}}\end{bmatrix}^{-1} \tag{59}$$

which has a computational complexity of $O(n(k+p)^2+(k+p)^3)$ in terms of matrix computation and requires k forward model runs. In most cases, due to the small truncation number used for the principal components, $n \gg k$. Thus, the dominant complexity term for the above equation is $O(n(k+p)^2)$. Each conditional realization can be computed as:

$$\mathbf{s} = [\mathbf{V}_k, \mathbf{X}]\left(\begin{bmatrix}\hat{\mathbf{a}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} + \mathbf{G}^T\mathbf{r}\right) \tag{60}$$

where $\mathbf{G}$ is the Cholesky factor of $\mathbf{C}_{\text{posteior}}$ satisfying $\mathbf{G}^T\mathbf{G} = \mathbf{C}_{\text{posterior}}$, and $\mathbf{r}$ is a sample from $N(\mathbf{0}, \mathbf{I}_{k+p})$. The computational complexity of the Cholesky decomposition is $O(k^3)$. Thus, to generate N_c conditional realizations for uncertainty analysis, the dominant computational complexity is approximately $O(n(k+p)^2 + (k+p)^3 + (k+p)^2 N_C + m(k+p)N_c)$ in respect of matrix computation, and k forward model runs.

5.4 Relation to Textbook GA

The textbook GA presents two formulas, named as the $\boldsymbol{\xi}$ form and the $\boldsymbol{\Lambda}$ form, respectively, in a line of P.K. Kitanidis's publications. In the following, we illustrate how the derived solutions on principal components are related to these two formulas based on the linear model.

5.4.1 The $\boldsymbol{\xi}$ Form

The first normal equation with respect to $\mathbf{a}$ in Equation (48) can be transformed to:

$$\hat{\mathbf{a}} = (\mathbf{H}\mathbf{V}_k)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{61}$$

where $\boldsymbol{\Sigma} = \mathbf{H}\mathbf{V}_k{\mathbf{V}_k}^T\mathbf{H}^T + \mathbf{R}$ is the $n \times n$ covariance matrix of $\mathbf{y}$ approximated by the low-rank $\mathbf{C}$. Introducing the $n \times 1$ vector of $\boldsymbol{\xi}$ through

$$\mathbf{y} - \mathbf{H}\mathbf{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}\boldsymbol{\xi} \tag{62}$$

Then, the equation for $\hat{\mathbf{a}}$ becomes:

$$\hat{\mathbf{a}} = \mathbf{V}_k{}^T \mathbf{H}^T \boldsymbol{\xi} \tag{63}$$

Substituting the above equation into Equation (48), we have:

$$(\mathbf{HX})^T \boldsymbol{\xi} = \mathbf{0} \tag{64}$$

Combining the above equations, we obtain:

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{HX} \\ (\mathbf{HX})^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \tag{65}$$

which is exactly the same as the $\boldsymbol{\xi}$ form of the textbook GA approach with $n + p$ equations. The only difference is that $\boldsymbol{\Sigma}$ is the low-rank approximation. Thus, the new solution based on principal components and the $\boldsymbol{\xi}$ form yield the same estimates, which are connected by Equation (63).

5.4.2 The $\boldsymbol{\Lambda}$ or Cokriging Form

The $\boldsymbol{\Lambda}$ form is also known as the cokriging formula, in which the unknown estimates are expressed as a linear combination of observations. Unlike the textbook approach, in which the unknown is the random field, $\hat{\mathbf{s}}$, our unknowns are principal component coefficients, $\hat{\mathbf{a}}$. Thus, we have:

$$\hat{\mathbf{a}} = \boldsymbol{\Lambda} \mathbf{y} \tag{66}$$

where $\mathbf{\Lambda} \in \mathbb{R}^{k \times n}$ is the coefficient matrix, also known as kriging coefficients. Substituting the above equation into Equation (48) yields:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{67}$$

By substituting the above equation into Equation (61), we obtain:

$$\hat{\mathbf{a}} = (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{y} - (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{68}$$

Thus, we have:

$$\mathbf{\Lambda} = (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1} - (\mathbf{H}\mathbf{V}_{\mathrm{k}})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1} \tag{69}$$

Define the $p \times K$ matrix $\mathbf{M}$ from the following expression:

$$\mathbf{M}^{\mathrm{T}} = (\mathbf{H}\mathbf{Z})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-\mathbf{1}} \tag{70}$$

Thus, the expression of $\mathbf{\Lambda}$ becomes:

$$\boldsymbol{\Sigma}\mathbf{\Lambda}^{\mathrm{T}} + \mathbf{H}\mathbf{X}\mathbf{M} = \mathbf{H}\mathbf{V}_{\mathrm{k}} \tag{71}$$

In addition, the expression of $\mathbf{\Lambda}$, we have:

$$\mathbf{\Lambda}\mathbf{H}\mathbf{X} = \mathbf{0} \tag{72}$$

Combing the above two equations yields the cokriging formula:

$$\begin{bmatrix} \mathbf{\Sigma} & \mathbf{HX} \\ (\mathbf{HX})^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{\mathrm{T}} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HV}_{\mathrm{k}} \\ \mathbf{0} \end{bmatrix} \tag{73}$$

We shall notice that $\mathbf{HV}_{\mathrm{k}}$ is the cross-covariance between $\mathbf{y}$ and $\mathbf{a}$. Compared with the classical cokriging method directly working on $\hat{\mathbf{s}}$, aside from the low-rank approximation of the covariance matrices, there are two additional differences. Firstly, the unbiasedness constraints are to ensure the known mean of $\mathbf{a}$ instead of the drift function, which is separated from $\mathbf{a}$, so that the mean of $\mathbf{s}$ is described by the drift function. Secondly, in the textbook cokriging method and the recently developed approach using low-rank approximation of covariance matrices [*Lee and Kitanidis*, 2014], the dimension of the kriging coefficient matrix $\mathbf{\Lambda}$ is $m \times n$, i.e., the kriging coefficients need to be computed for each $\hat{\mathbf{s}}$. However, to estimate $\hat{\mathbf{a}}$, the dimension of $\mathbf{\Lambda}$ is only $k \times n$, which is a much smaller matrix for large-dimensional inverse problems. In summary, we demonstrate that the estimation of principal component coefficients, $\mathbf{a}$, can be formulated through classical geostatistical approaches. However, direct inversion of $\hat{\mathbf{a}}$ only requires a linear system with $k + p$ equations, while the textbook geostatistical approaches need to solve $n + p$ equations, dominated by the number of observations. Moreover, for quasi-linear inverse problems, the Jacobian matrix in terms of the principal component coefficients can be directly used in the normal equations and there is no need for matrix multiplication to evaluate $\bar{\mathbf{H}}\mathbf{V}_{\mathrm{k}}$ and $\bar{\mathbf{H}}\mathbf{X}$. The total forward model runs for each iteration is reduced to $k + p + 1$.

5.5 Numerical Experiments

Two numerical experiments of two-dimensional steady-state hydraulic tomography are presented to demonstrate the validity of RGA. Table 5 lists the geostatistical and geometric parameters of the random field and the hydraulic parameters. Two heterogeneous fields of logarithmic hydraulic conductivity having different parameters are presented for validation of the proposed approach. The grid resolution is 1024×1024 with a total of 1.05 million unknowns. Sequential pumping tests are conducted in a well network consisting of 25 wells uniformly distributed in the domain. Top and bottom boundaries of the domain are impermeable, and left and right boundaries are constant heads with a specified natural gradient. During each pumping test, one well serves as the pumping well, and all the other 24 wells serve as monitoring wells. Thus, we obtain 600 observations of steady-state hydraulic heads in total.

Table 5 Geostatistical and hydrological parameters for the synthetic inverse problems.

Geostatistical Properties		
Domain Scale	$100\text{m} \times 100\text{m}$	$100\text{m} \times 100\text{m}$
Field Type	Logarithmic Field	Logarithmic Field
Resolution	1024×1024	1024×1024
Covariance Model	Gaussian model	Exponential model
Mean	$\mu = -5\text{m/s}$	$\mu = -5\text{m/s}$
Variance	$\sigma^2 = 4\text{m}^2/\text{s}^2$	$\sigma^2 = 6\text{m}^2/\text{s}^2$
Correlation length	$\text{lx} = 40\text{m}, \text{ly} = 20\text{m}$	$\text{lx} = 20\text{m}, \text{ly} = 40\text{m}$
Hydraulic Parameters		
Pumping Rate	$\text{Q} = 0.0075\text{m}^3/\text{s}$	$\text{Q} = 0.0075\text{m}^3/\text{s}$
Governing Equation	$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = 0$	
Top	Impermeable	Impermeable
Bottom	Impermeable	Impermeable
Left	$\text{h} = 0\text{m}$	$\text{h} = 0\text{m}$
Right	$\text{h} = 0\text{m}$	$\text{h} = 0\text{m}$

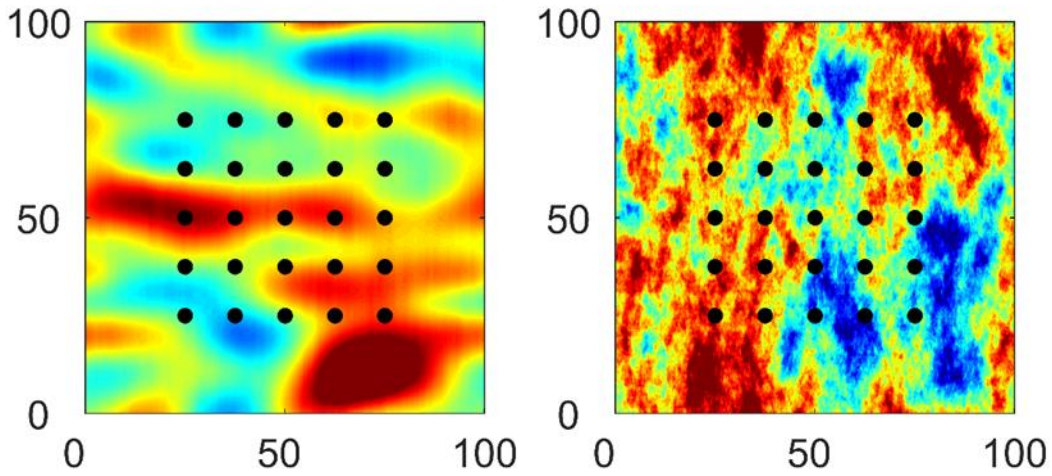

Figure 12 Random fields for two numerical experiments on 1024×1024 resolution. Field with Gaussian covariance function on left, field with exponential function on right.

The number of retained principal components in the inversion is 50 for both models of covariance function. Though the counted percentage of variance does not necessarily surpass 90% of the total variance, the results still indicate the effectiveness of the proposed approach as shown in Figure 13 and Figure 14. The inversion results for both cases are analyzed from the perspectives of best estimate and uncertainty of the field, best estimate and uncertainty of the principal component coefficients, and the reproduction of the available measurements. RGA yields satisfactory results in both cases. Both true fields are well represented by the best estimates. From a quantitative perspective, we adopt mapping accuracy as a criterion for evaluation of inversion precision [Kang et al., 2017]. If we define 10% of deviation as the criterion, the mapping accuracies of both cases are around 80%. Variances are generally low in the region enclosed by the existing well network, where the density of data information is high, and relatively high at the boundaries, where no measurement is taken. Subfigure Ds are unique for RGA, comparing the true principal component coefficients with the estimated coefficients. We shall see that for both cases, the true principal component coefficients are well contained in the domain formed by the upper and lower uncertainty bounds. Besides, scatters of simulated measurements against the available measurements also land on the 45^{o} lines in the coordinate, indicating a close fitting between these two. The inversion results demonstrates the ability of RGA in inverting large-scale, highly-heterogeneous conductivity fields, and justifies the use of the truncated principal components even though the total variance cannot be fully represented under such circumstances.

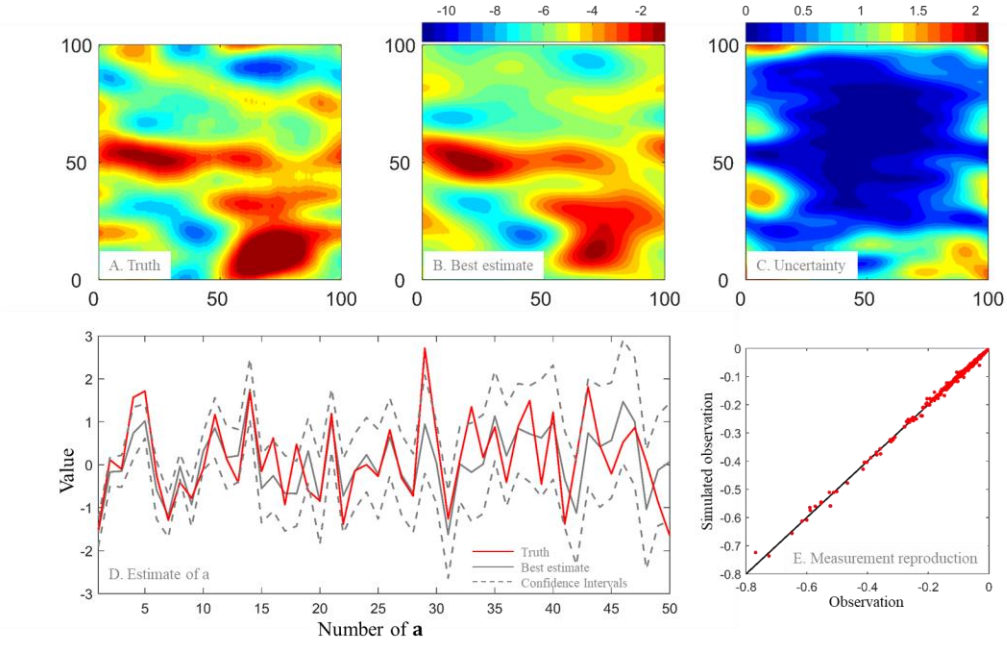

Figure 13 Inversion results for the case with Gaussian covariance function.

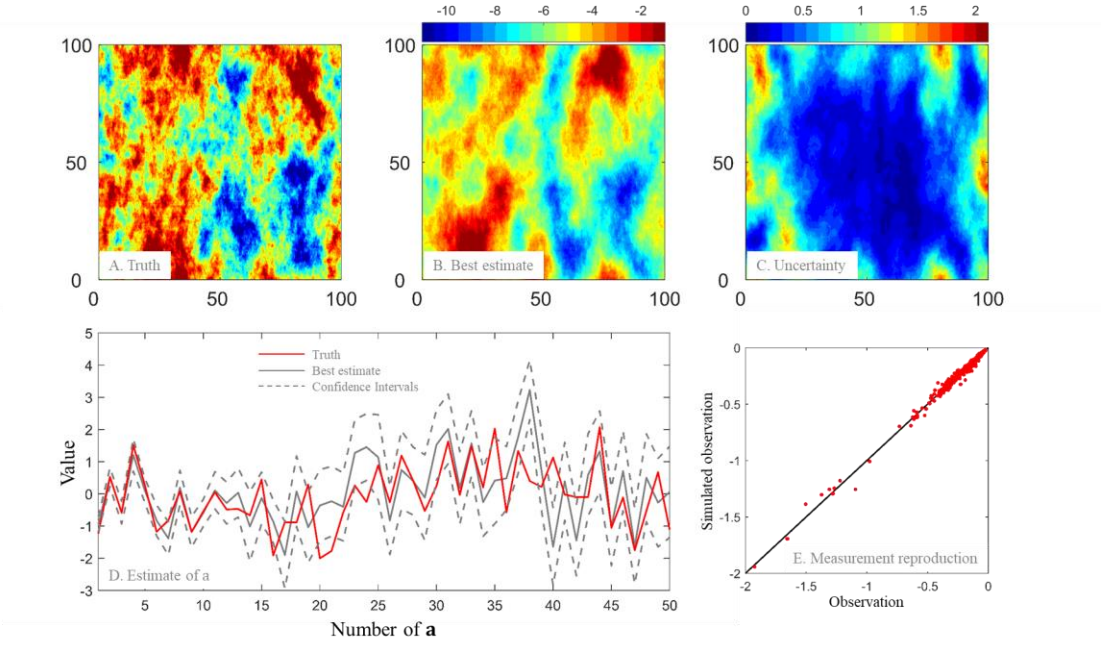

Figure 14 Inversion results for the case with exponential covariance function.

5.6 Conclusion

In this chapter, we reformulated the geostatistical inverse approach by estimating the principal component coefficients instead of the unknown parameter fields. We name the approach as reformulated geostatistical approach (RGA). Problem setup and advanced computational techniques are unified in the new framework. The normal equation system to be solved are reduced to $k + p$ equations, which are independent of the number of observations. The forward model runs for each iteration to determine the Jacobian matrix for nonlinear problems are reduced to $k + p + 1$ times. The so-called 'matrix-free' approach is now naturally built in RGA with less matrix multiplication to construct the normal equations. Thus, RGA is more efficient and scalable for high-dimensional inverse problems and problems with massive amount of observations. The application of principal component coefficients in inversion also raises a more intuitive approach to generate approximated posterior samples.

In addition, the specific application of hydraulic tomography indicates that a lower-rank truncation of the principal components are more advantageous even though the total variations of the original fields are under-represented, particularly for less smooth fields with an exponential covariance function.

CHAPTER 6. UPSCALING PRINCIPAL COMPONENT GEOSTATISTICAL APPROACH (UPCGA)

6.1 Introduction

Comprehension and modeling of flow and transport behavior in groundwater systems require accurate and efficient characterization of the underlying spatially distributed hydraulic conductivity field. Local-scale estimates of hydraulic conductivities can be obtained by conducting laboratory experiments on collected core samples from drilled boreholes [*Klute and Dinauer,* 1986a; *Zha et al.,* 2018]. However, high costs of well drilling hinder this approach from mapping field-scale hydraulic conductivity distributions onto a fine resolution. Therefore, field-scale characterization is usually achieved by inverse approaches to infer hydraulic conductivities from measurements of aquifer tests, such as pumping tests, slug tests, constant-head tests, and tracer tests etc. [*Cardiff et al.,* 2009; *Cirpka et al.,* 2007; *Fienen et al.,* 2006; *Liao and Cirpka,* 2011; *Yeh and Liu,* 2000; *Zhang et al.,* 2014; *Zhao et al.,* 2018]. Such inverse problems can be conveniently addressed in a Bayesian framework, which formulate the posterior distribution of unknown variables constrained by the likelihood function for data fitting and the prior distribution that is typically described by spatial correlation functions of hydraulic conductivities [*Gavalas et al.,* 1976; *Kitanidis,* 1997; *Neuman,* 1980; *Woodbury and Rubin,* 2000].

Various numerical approaches have been developed to solve Bayesian inverse problems. Ensemble-based approaches include Markov chain Monte Carlo (MCMC) that simulates a large number of the posterior samples by running a Markov chain [*Cui et al.,*

2014; *Geman and Geman,* 1984a; *Hastings,* 1970; *Martin et al.,* 2012; *Vrugt et al.*, 2009], ensemble Kalman filter and ensemble smoother approaches that simulates samples from a Gaussian approximation of the posterior distribution [*Zhang et al.,* 2018; *Zhou et al.*, 2012]. In this chapter, our focus is on the geostatistical approach (GA), which is an objective method and generally requires a smaller computational budget than ensemble-based methods [*Hoeksema and Kitanidis*, 1984; *Kitanidis and Vomvoris*, 1983a; *Kitanidis,* 1995]. However, for highly-parameterized Bayesian inverse problems that admit high-resolution hydraulic conductivity fields, the geostatistical approach still requires efficient implementations to facilitate the computation [*Lee and Kitanidis,* 2014; *Liu and Kitanidis*, 2011]. The reasons lie in two aspects: numerical solver needs to be performed on a high-resolution grid for multiple times in nonlinear problems [*Kitanidis,* 1995]; and the large number of unknowns increases the cost of matrix computation during the inversion.

Many efforts have been dedicated to reduce the computational cost of the geostatistical approach. *Nowak et al.* [2003] proposed an efficient approach of computing the cross-covariance matrix based on circulant embedding and the fast Fourier transform, which was successfully applied to a sandbox inverse problem [*Nowak and Cirpka*, 2006a]. *Liu and Kitanidis* [2011] proposed a sparse representation of the prior covariance matrix that facilitates the storage and the associated computation, and proved its effectiveness in an inverse problem of sandbox hydraulic tomography. For large-scale linear geostatistical inverse problems, *Ambikasaran et al.* [2013] and *Saibaba et al.* [2012] proposed hierarchical matrices to speed up the solving process. For geostatistical inverse problems with huge volume of observational data, *Klein et al.* [2017] proposed a preconditioned conjugate gradient algorithm, and *Lin et al.* [2017] proposed a randomized data reduction

algorithm. For inversion of pumping and tracer tests, temporal moments was applied to transform transient forward models to steady-state models for faster model evaluations [*Li et al.,* 2005; *Nowak and Cirpka,* 2006a; *Pollock and Cirpka,* 2008; *Yin and Illman,* 2009; *Zhu and Yeh,* 2006].

Recently, *Kitanidis and Lee* [2014] introduced the principal component analysis (PCA) in the quasi-linear geostatistical approach to cut down the computational cost of Bayesian inverse problems, named as principal component geostatistical approach (PCGA), and demonstrated its effectiveness and efficiency in several applications [*Kang et al.,* 2017; *Lee and Kitanidis,* 2014; *Lee et al.,* 2016]. PCGA decreases the computational overhead associated with the covariance matrix and improves the computational efficiency for the Jacobian matrices by a‘matrix-free’ approach. A similar idea was also incorporated into the successive co-kriging estimator for hydraulic tomography applications [*Zha et al.,* 2018].

In CHAPTER 5, we have reformulated the PCGA for a more compact normal equation that provides direct solution to the principal component coefficients. This reformulated geostatistical approach (RGA) also stimulates an upscaling approach by having principal component coefficients acting as a group of latent variables, a similar idea accepted in CHAPTER 4. Therefore, in this chapter, we develop a new geostatistical approach based on RGA and the upscaling approach that solves Bayesian inverse problems only by running the coarse-grid forward model. The new approach is named as the upscaling principal component geostatistical approach (UPCGA). We have also noticed that the idea of running forward models on coarse-grid fields was applied by *Efendiev et*

al. [2005] in a two-stage MCMC implementation. However, the coarse-grid model was mainly employed for eliminating highly biased samples in their implementation.

6.2 Upscaling Principal Component Geostatistical Approach (UPCGA)

The UPCGA approach inherits the idea from PCGA and RGA to exploit the low-rank structure of the $\mathbf{C}$ matrix. However, instead of inferring a fine-resolution random field $\mathbf{s}_1$ (which is denoted as $\mathbf{s}$ in CHAPTER 5), UPCGA infers a group of latent variable by merely running the forward model on an upscaled coarse-resolution field, $\mathbf{s}_2$. In the rest of this chapter, $\mathbf{s}_1$ is used to refer to the fine-resolution random field while $\mathbf{s}_2$ is used to represent the upscaled, coarse-resolution counterpart of it.

Using Equation (41) to express the linear expansion of $\mathbf{s}_1$ with respect to the principal component coefficients yields

$$\mathbf{s}_1 \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{V}_\mathrm{k}\mathbf{a} \tag{74}$$

where $\mathbf{a}$ is a standard multi-Gaussian random vector, and $\mathbf{V}_\mathrm{k} = [\mathbf{v_1}\ \mathbf{v_2}\ \mathbf{v_3}\ \ldots\ \ldots\ \mathbf{v_k}]$. That is, the unknown variable $\mathbf{s}$ can be constructed via the above equation once the principal component coefficients $\mathbf{a}$ are inverted. The MAP estimate of $\mathbf{a}$ can be found out by the quasi-linear approach, which formulates a $(\mathrm{k}+\mathrm{p}) \times (\mathrm{k}+\mathrm{p})$ linear system as:

$$\begin{bmatrix} \bar{\mathbf{H}}_\mathbf{a}^\mathrm{T}\mathbf{R}^{-1}\bar{\mathbf{H}}_\mathbf{a} + \mathbf{I} & \bar{\mathbf{H}}_\mathbf{a}^\mathrm{T}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^\mathrm{T}\mathbf{R}^{-1}\bar{\mathbf{H}}_\mathbf{a} & \bar{\mathbf{H}}_{\boldsymbol{\beta}}^\mathrm{T}\mathbf{R}^{-1}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_\mathbf{a}^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_\mathbf{a}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_\mathbf{a}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \end{bmatrix} \tag{75}$$

and the updated $\mathbf{s}_1$ can be constructed as

$$\mathbf{s}_1 = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{V}_k\hat{\mathbf{a}} \tag{76}$$

In cases where $\boldsymbol{\epsilon}$ is from an i.i.d. multi-Gaussian distribution, $\mathbf{R} = \sigma^2\mathbf{I}$, the linear system becomes

$$\begin{bmatrix} \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\bar{\mathbf{H}}_{\mathbf{a}} + \sigma^2\mathbf{I} & \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\bar{\mathbf{H}}_{\mathbf{a}} & \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\bar{\mathbf{H}}_{\boldsymbol{\beta}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{\mathbf{a}}^{\mathrm{T}}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \\ \bar{\mathbf{H}}_{\boldsymbol{\beta}}^{\mathrm{T}}\left(\mathbf{y} - \mathbf{f}\left(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}\right) + \bar{\mathbf{H}}_{\mathbf{a}}\bar{\mathbf{a}} + \bar{\mathbf{H}}_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\right) \end{bmatrix} \tag{77}$$

which takes $O(npk + p^2n)$ more matrix multiplications than the traditional formulation in PCGA. For cases with $p = 1$, i.e., a constant mean, the extra computational cost is manageable. Additionally, the above linear system is supposed to have better scalability, as its dimension is independent of the problem size or the observation data size.

However, solving Equation (77) iteratively still requires a substantial number of forward model runs on the fine-resolution grid. To lower the computational cost of each forward model evaluation, the original field on the fine-resolution grid is upscaled to be a coarse-resolution field. Assuming there is an upscaling matrix $\mathbf{U}$ that is applicable to transform the original field $\mathbf{s}_1$:

$$\mathbf{s}_2 = \mathbf{U}\mathbf{s}_1 = \mathbf{U}\mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{V}_k\mathbf{a} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{V}_k'\mathbf{a} \tag{78}$$

where $\mathbf{s}_2 \in \mathbb{R}^{m_2 \times 1}$ and m_2 is a much smaller number than m_1. The simplest upscaling matrix $\mathbf{U}$ for uniform grids is the spatial average matrix, which is also very computationally efficient as a sparse matrix. Besides, it also maintains a sound physical meaning in

applications of hydraulic conductivity, as the spatial average of the natural log conductivity corresponds to the geometric mean of the conductivity. By applying the upscaling matrix, the MAP estimate of $\mathbf{a}$ can also be obtained with $\mathbf{s}_2$. The new observation equation becomes:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}_2) + \boldsymbol{\epsilon}_2 \tag{79}$$

where $\boldsymbol{\epsilon}_2$ represents errors raised by using $\mathbf{s}_2$ instead of $\mathbf{s}_1$ in forward model evaluations plus the original error term $\boldsymbol{\epsilon}$. Since $\mathbf{s}_2$ is the upscaled field, the epistemic error, $\boldsymbol{\epsilon}_2$, also includes the differences between the simulation results of the coarse-resolution and fine-resolution forward models. For simplicity, we assume $\boldsymbol{\epsilon}_2$ to be a multi-Gaussian distribution $\mathrm{N}(\mathbf{0}, \mathbf{R}')$. In this chapter, the correlation matrix $\mathbf{R}'$ is assumed to be $\sigma'^2\mathbf{I}$. The validity and applicability of this assumption will be further investigated in the following numerical experiments. Thus, the formulated posterior distribution is

$$\begin{aligned} \mathrm{p}(\mathbf{a}, \boldsymbol{\beta}|\mathbf{y}) \propto \exp\Big(&-\frac{1}{2}\big(\mathbf{y} - \mathbf{f}(\mathbf{X}'\boldsymbol{\beta} + \mathbf{V}_\mathrm{k}'\mathbf{a})\big)^\mathrm{T}\mathbf{R}'^{-1}\big(\mathbf{y} - \mathbf{f}(\mathbf{X}'\boldsymbol{\beta} + \mathbf{V}_\mathrm{k}'\mathbf{a})\big) \\ &-\frac{1}{2}\mathbf{a}^\mathrm{T}\mathbf{a}\Big) \end{aligned} \tag{80}$$

Accordingly, the linear system is revised to be

$$\begin{aligned} &\begin{bmatrix} \bar{\mathbf{H}}'^\mathrm{T}_\mathbf{a}\mathbf{R}'^{-1}\bar{\mathbf{H}}'_\mathbf{a} + \mathbf{I} & \bar{\mathbf{H}}'^\mathrm{T}_\mathbf{a}\mathbf{R}'^{-1}\bar{\mathbf{H}}'_{\boldsymbol{\beta}} \\ \bar{\mathbf{H}}'^\mathrm{T}_{\boldsymbol{\beta}}\mathbf{R}'^{-1}\bar{\mathbf{H}}'_\mathbf{a} & \bar{\mathbf{H}}'^\mathrm{T}_{\boldsymbol{\beta}}\mathbf{R}'^{-1}\bar{\mathbf{H}}'_{\boldsymbol{\beta}} \end{bmatrix}\begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} \\ &\qquad = \begin{bmatrix} \bar{\mathbf{H}}'^\mathrm{T}_\mathbf{a}\mathbf{R}'^{-1}\big(\mathbf{y} - \mathbf{f}(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}) + \bar{\mathbf{H}}'_\mathbf{a}\bar{\mathbf{a}} + \bar{\mathbf{H}}'_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\big) \\ \bar{\mathbf{H}}'^\mathrm{T}_{\boldsymbol{\beta}}\mathbf{R}'^{-1}\big(\mathbf{y} - \mathbf{f}(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}) + \bar{\mathbf{H}}'_\mathbf{a}\bar{\mathbf{a}} + \bar{\mathbf{H}}'_{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}\big) \end{bmatrix} \end{aligned} \tag{81}$$

where $\overline{\mathbf{H}}'_{\mathbf{a}}$ and $\overline{\mathbf{H}'}_{\boldsymbol{\beta}}$ is computed based on $\mathbf{V}_{\mathrm{k}}'$ and $\mathbf{X}'$. It worth noting that formulating or solving the above linear system only involves with forward evaluations $\mathbf{f}(\mathbf{s}_2)$, which is much faster than the original, computationally heavy forward model $\mathbf{f}(\mathbf{s}_1)$. $\mathbf{s}_2$ between iterations can be updated by

$$\mathbf{s}_2 = \mathbf{X}'\hat{\boldsymbol{\beta}} + \mathbf{V}_{\mathrm{k}}'\hat{\mathbf{a}} \tag{82}$$

which is achieved by only running the coarse-resolution forward model, and once the algorithm converges, the original fine-resolution solution $\mathbf{s}_1$ can be recovered by Equation (76). It is worth noting that by using different $\mathbf{V}_{\mathrm{k}}'$, or say, by using different upscaling matrix $\mathbf{U}$, 'inversion results' on different resolutions can be constructed while they share the same group of underlying principal component coefficients. However, such 'inversion results' are just estimates of the spatial averages of the solution on the original fine resolution. An example and a detailed explanation will be presented in the numerical experiment section of this chapter.

Besides a more efficient process of computing MAP estimate, the upscaling approach can also provide faster uncertainty quantifications. After the iteration procedure has converged, the posterior covariance of $\mathbf{s}_1$ can be computed as:

$$\begin{bmatrix}\mathbf{V}_{\mathrm{k}}{}^{\mathrm{T}}\\ \mathbf{X}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\overline{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a}}\mathbf{R}'^{-1}\overline{\mathbf{H}}'_{\mathbf{a}}+\mathbf{I} & \overline{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a}}\mathbf{R}'^{-1}\overline{\mathbf{H}}'_{\boldsymbol{\beta}}\\ \overline{\mathbf{H}}'^{\mathrm{T}}_{\boldsymbol{\beta}}\mathbf{R}'^{-1}\overline{\mathbf{H}}'_{\mathbf{a}} & \overline{\mathbf{H}}'^{\mathrm{T}}_{\boldsymbol{\beta}}\mathbf{R}'^{-1}\overline{\mathbf{H}}'_{\boldsymbol{\beta}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{V}_{\mathrm{k}}{}^{\mathrm{T}}\\ \mathbf{X}^{\mathrm{T}}\end{bmatrix} \tag{83}$$

We shall notice that in the above equation, the covariance matrix of the principal component coefficients is computed on the coarse-grid resolution, while the covariance

matrix of the original field $\mathbf{s}_1$ is recovered by multiplication of $\begin{bmatrix}\mathbf{V}_k^T \\ \mathbf{X}^T\end{bmatrix}$ and its transpose on both sides. Therefore, compared with the uncertainty quantification procedures in RGA, the proposed approach quantifies the posterior uncertainty only requests coarse-grid forward model $\mathbf{f}(\mathbf{s}_2)$, which further strengthens the computational advantage of the upscaling approach.

The covariance matrix of $\begin{bmatrix}\hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}}\end{bmatrix}$ can also be employed to compute conditional realizations from a Gaussian approximation of the posterior. It follows the same procedure as what has been shown in CHAPTER 5. However, the computational overhead can be alleviated further because of the use of coarse-grid forward models.

6.3 Numerical Experiments

The effectiveness and computational cost of the proposed UPCGA is investigated in three large-scale hydraulic tomography problems with each one addressing a different concern regarding this upscaling approach [*Gottlieb and Dietrich*, 1995; *Liu et al.*, 2002; *Yeh and Liu*, 2000]. The first case is a multi-Gaussian random field; the second case is a structured heterogeneous field, while both cases are inverted by steady-state pumping tests. The third case involves with the inversion of another multi-Gaussian random field, and it is inverted by a transient pumping test.

All of the hydraulic conductivity fields inverted in this section have a fine resolution of 512×512, which yields 262,144 unknowns to be estimated in the inverse problem. For the first two cases, both UPCGA and PCGA are applied to estimate the underlying hydraulic conductivity fields. PCGA directly implements the inversion on the fine

resolution, while UPCGA applies different upscaling resolutions, 128×128, 64×64, 32×32 and 16×16, to compare the performance and examine the potentially lowest resolution necessary for producing satisfactory inversion results. The third case only scrutinizes the applicability of UPCGA for transient pumping case and contrasts the inversion results to those obtained by steady-state pumping data.

All numerical experiments are implemented on a desktop computer equipped with Intel® Xeon® W2102 CPU @ 2.90 GHz 2.90 GHz processor and 8.00 GB RAM.

6.3.1 Case1. Inversion of a Multi-Gaussian Random Field via Steady-State Sequential Pumping Tests

6.3.1.1 Inverse Problem Settings

In the first numerical experiment, a multi-Gaussian random field is generated as the true field of logarithmic hydraulic conductivities, as shown in Figure 15. Table 6 summarizes the geostatistical parameters, hydraulic parameters, and the governing equation and associated boundary conditions. A well network with 25 wells, represented by black dots in Figure 15, are installed at the field. In each phase of the hydraulic tomography, one of the wells acts as the pumping well, and the steady-state hydraulic heads in the other wells are recorded as observational data. The total amount of available observational data is 600. We also consider the effect of errors of measurements by contaminating the synthetic data by adding 1.5%~3% normal random errors. In this case, the number of retained principal components for both PCGA and UPCGA are 30, which accounts for more than 95% of the total variance in the covariance matrix.

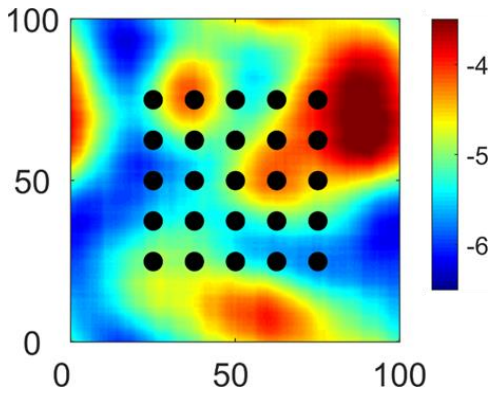

Figure 15 True random field with 512×512 resolution.

Table 6 Geostatistical and hydrological parameters for the synthetic inverse problem Case 1.

Geostatistical Properties	
Domain Scale	$l_x \times l_y = 100\text{m} \times 100\text{m}$
Field Type	Log 10 Field
Resolution	512×512
Covariance Model	Gaussian model
Mean	$\mu = -5\text{m/s}$
Variance	$\sigma^2 = 0.5\text{m}^2/\text{s}^2$
Correlation length	$l = 20\text{m}$
Hydraulic Parameters	
Pumping Rate	$Q = 200\text{m}^3/\text{d}$
Governing Equation	
$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = Q$$	
Top	Impermeable
Bottom	Impermeable
Left	$h = 0.1\text{m}$
Right	$h = 0\text{m}$

6.3.1.2 Results and Discussion

The inversion results are shown in Figure 16. The first row shows the best estimates of the logarithmic conductivity field while the second row represents the variances of the estimates. The case with forward model running on the 512×512 grid is solved by PCGA, while the rest are solved by UPCGA. Compared with the true field and the best estimate obtained by PCGA on the fine grid, the best estimates obtained by UPCGA successfully capture the main features of the field. In fact, no obvious quality difference can be observed between PCGA and UPCGA. However, the total computational time of the 16×16 case is only about 7s, while that of the 512×512 case is about 12,330s. The computational time is reduced up to a factor of 1820 by the upscaling approach.

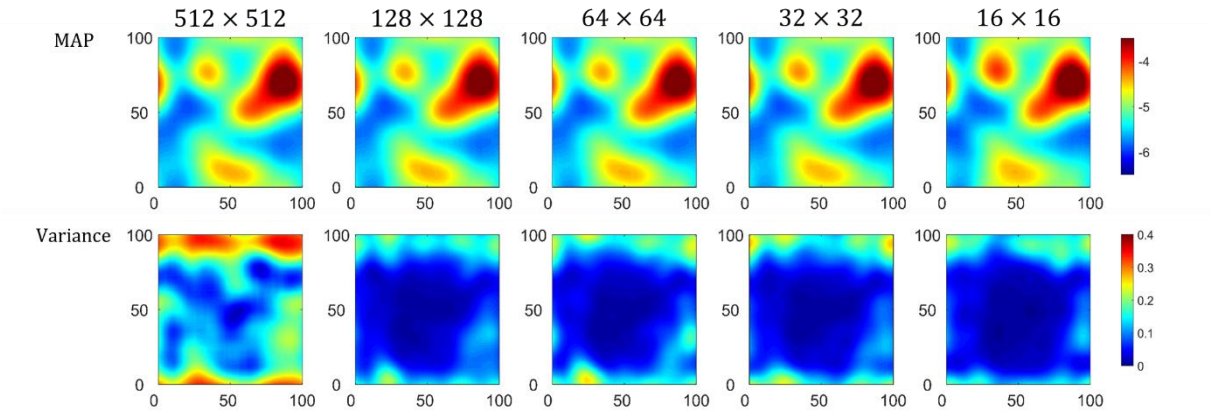

Figure 16 MAP estimates (the first row) and uncertainty quantifications (the second row) for Case 1. Solutions on 512×512 resolution is solved by PCGA.

The variance maps are computed following the procedures described in section 6.2. The variance map has much lower values around the locations of the wells, indicating that the estimate of the field has lower uncertainty around where the data are collected. It can also be observed that UPCGA produces variance maps with much lower values, which can be explained by the averaged principal components $\mathbf{V}_k'$. As when the upcaling matrix $\mathbf{U}$ is applied to the original principal components $\mathbf{V}_k$, a certain amount of high-frequency variations is weighted out due to the spatial average. That is, UPCGA yields much lower variance estimates than PCGA because of the loss in high-frequency noise.

To evaluate the computational performances of the proposed approach in a more statistically meaningful way, we generate five different random fields with the same geostatistical parameters for inversion. The computational time and the number of forward model runs for different upscaled resolutions are plotted in Figure 17. The computational time has no error bar, as the variance of the computational time is too small to be visible under the figure's scale.

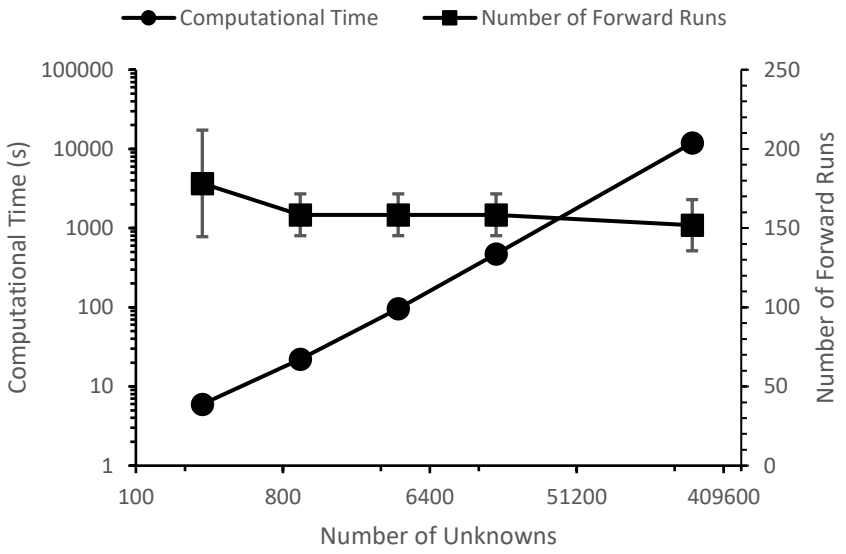

Figure 17 Computational costs for Case 1. Statistics of 512×512 resolution is based on results of PCGA.

The computational time approximately follows a linear relationship with the number of unknowns in the log-log plot, which implies a linear reduction of computational time by UPCGA. Moreover, the proposed approach does not jeopardize the convergence speed of PCGA. The numbers of forward model runs to convergence for all tests are between $132\sim232$. Even though the number of forward model runs increases slightly when the field resolution is upscaled to 16×16, its impact on the total computational time is negligible due to much faster forward model implementation.

The accuracy of UPCGA is further analyzed via two quantitative criteria: map accuracy (MA) (Kang et al., 2017) and mean percentage of absolute error (MPAE) of the logarithmic conductivity field. Map accuracy represents the decimal proportion of 'correctly' estimated conductivity values, where the 'correctness' is defined to be within an acceptable deviation from the true conductivity value. In this chapter, the tolerable deviation is set to be 15% of the absolute value between the maximum conductivity and the minimum conductivity of the true field [*Kang et al.*, 2017].

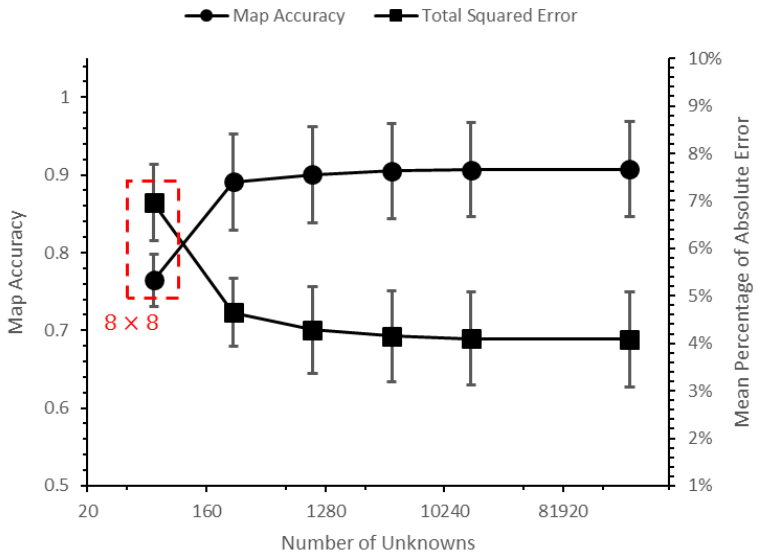

Figure 18 Results accuracy facts for Case 1. Statistics of 512×512 resolution is based on results of PCGA.

Figure 18 shows the curves of map accuracy and mean percentage of absolute error for forward models with different field resolutions. The results maintain a high level of accuracy, except for the case on 8×8 resolution, the coarsest grid used for case 1. This motivates a practical procedure of choosing the possibly coarsest resolution for UPCGA: we can start from the very coarsest resolution that retains the physical meaning of the forward model, which is 8×8 in this experiment, to a higher resolution by a certain enlarging factor until the estimated fields from two consecutive resolutions possess insignificant differences. This also justifies our use of 16×16 as the coarsest resolution for inversion, as UPCGA still yields high quality estimates of the fields with mean MA of 0.9 and MPAE of 5%. However, it is worth noting that the average performance of PCGA on 512×512 resolution is slightly better than that of UPCGA on 16×16. This inspires another potential implementation method that we could run UPCGA for the first few iterations on coarse-grid fields, which provides a quick convergence to MAP, and finish the inversion by one or two iterations on the fine-grid field, which yields better precisions on the final estimate.

6.3.2 Case 2. Inversion of a Structured Heterogeneous Field via Steady-State Sequential Pumping Tests

6.3.2.1 Inverse Problem Settings

In this numerical experiment, a structured heterogeneous hydraulic conductivity field (see Figure 19) is constructed to test the eligibility of UPCGA for non-Gaussian fields. The field structure is similar to the porous media packed in a sand box [*Liu and Kitanidis,* 2011], where several low-permeability inclusions are embedded in a homogeneous

medium. Table 7 summaries the hydraulic conductivities, hydraulic parameters and the boundary conditions. The pumping and monitoring setups are the same as those in the first experiment. However, there are ten more pumping wells are added. Thus, the number of available observational data increases to be 850 for this experiment. The observational data is also contaminated by 1.5%~3% normal random error. The solely pumping wells are represented by white dots in the field, and the dual-function wells are represented as black dots in the field.

As the structured field is not produced by any geostatistical covariance function, the geostatistical parameters for inversion needs to be specified. *Yeh and Liu* [2000] and *Liu et al.* [2007] discussed that the choice of σ^2 and L does not make significant differences on the inversion results when observational data is abundant. In this case, the geostatistical parameters are assigned as: $\sigma^2 = 2.5$, $\mathrm{L} = 20\mathrm{m}$, and $\mu = -1.76$. The number of retained principal components for both PCGA and UPCGA approaches are 50, which accounts for more than 95% of the total variance in the covariance matrix.

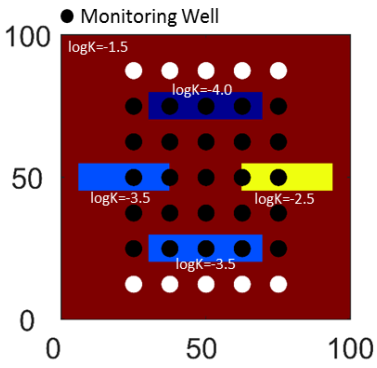

Figure 19 True structured field with 512×512 resolution.

Table 7 Parameters for the synthetic inverse problem Case 2.

Geostatistical Properties	
Field Type	Natural Log Field
Domain Scale	$l_x \times l_y = 100\text{m} \times 100\text{m}$
Resolution	512×512
Hydraulic Parameters	
Pumping Rate	$Q = 3.17\ \text{m}^3/\text{d}$
Governing Equation	
$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x}+\frac{\partial h}{\partial y}\right)\right)=Q$$	
Boundary Conditions	
Top	$h = 0\text{m}$
Bottom	$h = 0\text{m}$
Left	$h = 0.1\text{m}$
Right	$h = 0\text{m}$

6.3.2.2 Results and Discussion

The inversion results of the second case are shown in Figure 20. The MAP estimates of the log conductivity field are plotted in the first row while the variances are shown in the second row. The case with forward model running on 512×512 field is solved by PCGA, and the others are solved by UPCGA. Figure 20 indicates that UPCGA is capable of capturing the main features of the structured field even if the forward solver is upscaled to much coarser grids. Similar to the first experiment, the uncertainties of the best estimates of UPCGA are much lower than that of PCGA.

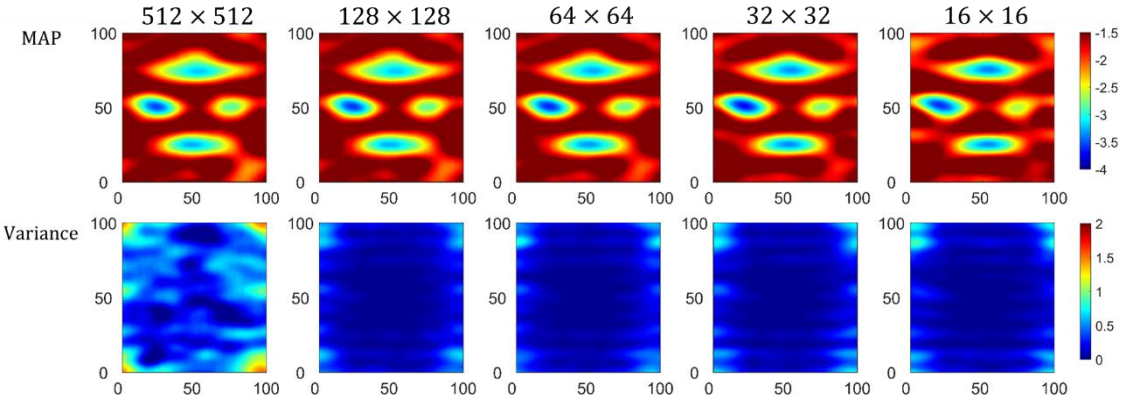

Figure 20 MAP estimates (the first row) and uncertainty quantifications (the second row) for Case 2. Solutions on 512×512 resolution is solved by PCGA.

Table 8 summarizes the computational cost and inversion accuracy. The results indicate that even though the prior information is highly biased, both PCGA and UPCGA provide inversion results with satisfactory MAs and MPAEs. Moreover, UPCGA on coarse grids demonstrates computational advantages over PCGA by significantly reducing the time consumption of each forward model run. That the performance is not as good as in the first experiment is mainly because of the sharp changes at the inclusion boundaries, which can hardly be described by any existing geostatistical covariance function.

Table 8 Computational performances for Case 2. Statistics of 512×512 resolution is based on results of PCGA.

	# of Forward Runs	Total Time (s)	Map Accuracy	Mean Percentage of Absolute Error
512×512	212	2.31E+04	0.65	19.77%
128×128	212	8.45E+02	0.66	19.50%
64×64	265	2.20E+02	0.68	19.36%
32×32	265	45.66267	0.68	19.78%
16×16	265	11.38905	0.60	21.78%

6.3.3 Case 3. Inversion of a Multi-Gaussian Random Field via Transient Sequential Pumping Tests

6.3.3.1 Inverse Problem Settings

In this numerical experiment, a multi-Gaussian random field with exponential covariance function is generated as the true field of logarithmic hydraulic conductivities (Figure 21). Table 9 contain parameters of the forward hydraulic model. A same well network composed of 25 evenly distributed wells is employed in this case. During the sequential pumping test, one of the wells acts as the pumping well, and the transient hydraulic heads in the other wells are recorded as observational data. The total amount of available observational data is 6000. We also consider the effect of errors of measurements by contaminating the synthetic data by adding $1.5\%\sim3\%$ normal random errors. In this case, we only investigate the performance of UPCGA in inverting transient pumping data, which is featured with inversion results on 64×64, 32×32 and 16×16. The number of retained principal components for this case is 50.

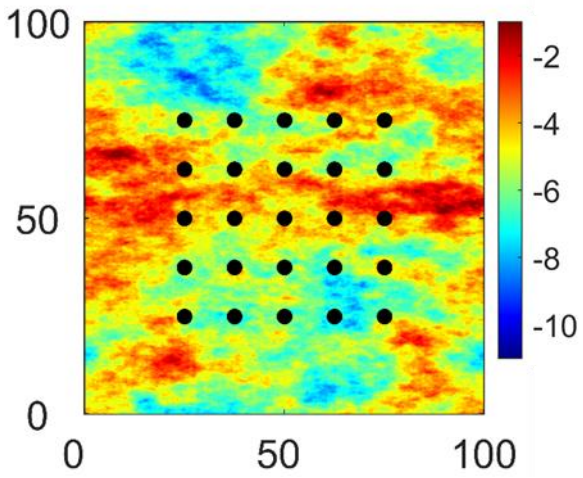

Figure 21 True structured field with 512×512 resolution (Exponential covariance function).

Table 9 Parameters for the synthetic inverse problem Case 3.

Geostatistical Properties	
Domain Scale	$l_x \times l_y = 100\text{m} \times 100\text{m}$
Field Type	Log 10 Field
Resolution	512×512
Covariance Model	Gaussian model
Mean	$\mu = -5\text{m/s}$
Variance	$\sigma^2 = 2\text{m}^2/\text{s}^2$
Correlation length	$l = 20\text{m}$
Hydraulic Parameters	
Pumping Rate	$Q = 0.0075\text{m}^3/\text{s}$
Storage Coefficient	$S = 10^{-4}$
Governing Equation	
$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = Q + \frac{\partial h}{\partial t}$$	
Top	Impermeable
Bottom	Impermeable
Left	$h = 0\text{m}$
Right	$h = 0\text{m}$

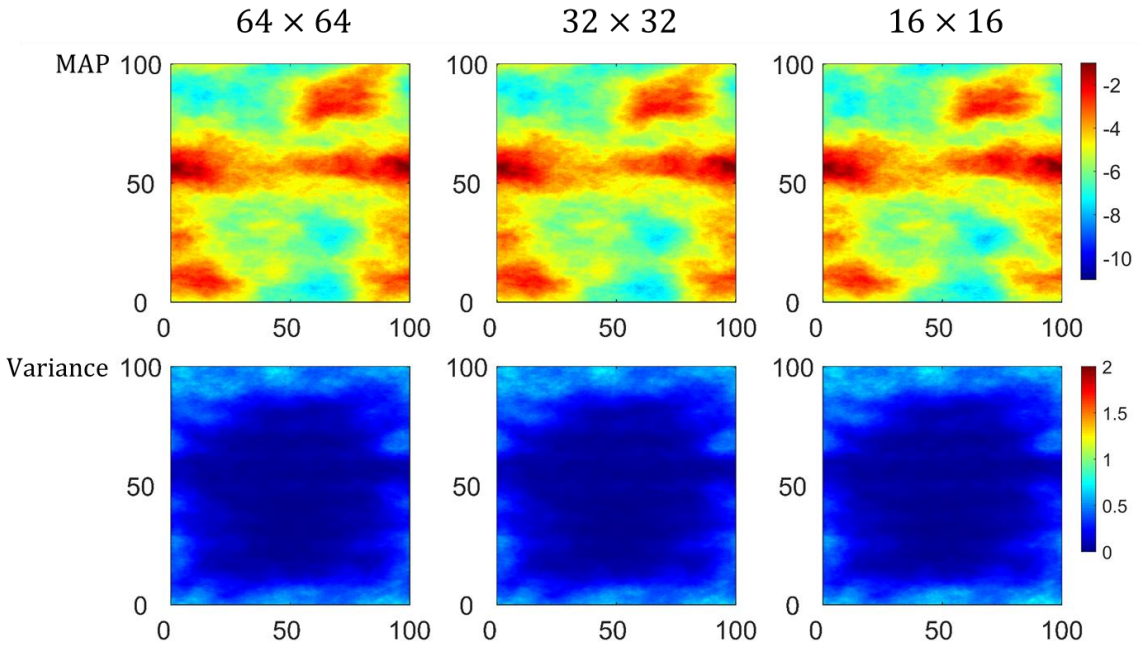

Figure 22 MAP estimates (the first row) and uncertainty quantifications (the second row) for Case 3.

6.3.3.2 Results and Discussion

The results in Figure 22 indicate that the proposed upscaling approach is also capable of inverting transient sequential pumping tests data, as it captures the main features of the underlying random field. Setting the threshold to be 10% of the difference between maximum and minimum in the original field, MAs of resolutions 64×64, 32×32 and 16×16 are 84.90%, 84.07% and 81.81% respectively, which are indicative of the success of the upscaling approach. By using transient pumping data, the magnitude of variances is also lowered as demonstrated in the uncertainty maps. The suppressed uncertainty is mostly provoked by a much larger data volume.

The reduction of computational overhead is also significant. Inversions on three different resolutions share a same number of required forward model runs, while, due to a decreasing running time for each forward model, the total computational times are 741s, 136s and 32.6s.

6.3.4 Discussion on Upscaling Effects

The spatial average of the logarithmic hydraulic conductivity values corresponds to the geometric mean of the actual hydraulic conductivity values, which has been proved sufficiently effective for hydraulic tomographic problems presented above. In this section, a detailed discussion regarding the upscaling effect is presented.

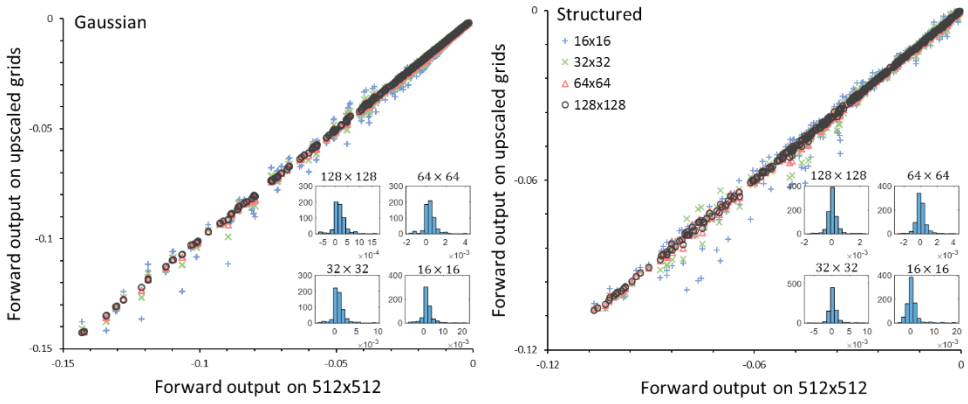

Figure 23 Scatter plots of the forward model outputs on the upscaled fields versus the forward model output on the original field, and the histograms of the deviations.

Figure 23 shows the scatter plots of the forward model outputs using different upscaled resolution against the forward model output on the original one. That the scatter points roughly fall on the diagonal line indicates an agreement of forward model outputs across different field resolutions, but as the grid is upscaled further, there are more points diverted off the diagonal line. The errors of simulation are also illustrated by histograms, which do not show any obvious skewness. That also helps to explain why a Gaussian assumption of the error term $\boldsymbol{\epsilon}_2$ yields comparable inversion results in these two cases. However, when these deviations such as numerical errors suppress the real trend of the forward model output, special treatment is needed [*Köpke et al.,* 2019].

As stated before, by employing different $\mathbf{V}_{\mathrm{k}}'$, we can recover the spatial averages of the original field on different coarse resolutions. This hypothesis is supported by the two leftmost columns in Figure 24, which represents the 16×16 spatially averaged true field and the 16×16 field constructed by the principal component coefficients. Subplots in the first column and the second column almost share the same spatial pattern, and if are treated as the true fields and the best estimates, MA and MPAE of the Gaussian case are 0.91 and 4%, while those of the structured case are 0.60 and 16%. On the other hand, a direct inversion on resolution 16×16, the rightmost column, deviates more from the averaged true field, characterized by an MA of 0.84 and an MPAE of 5% for the Gaussian case, and an MA of 0.60 and an MPAE of 16% for the structured case. It implies that the proposed approach is different from a direct concatenation of an inversion of a coarse field and a downscaling of it to a fine resolution.

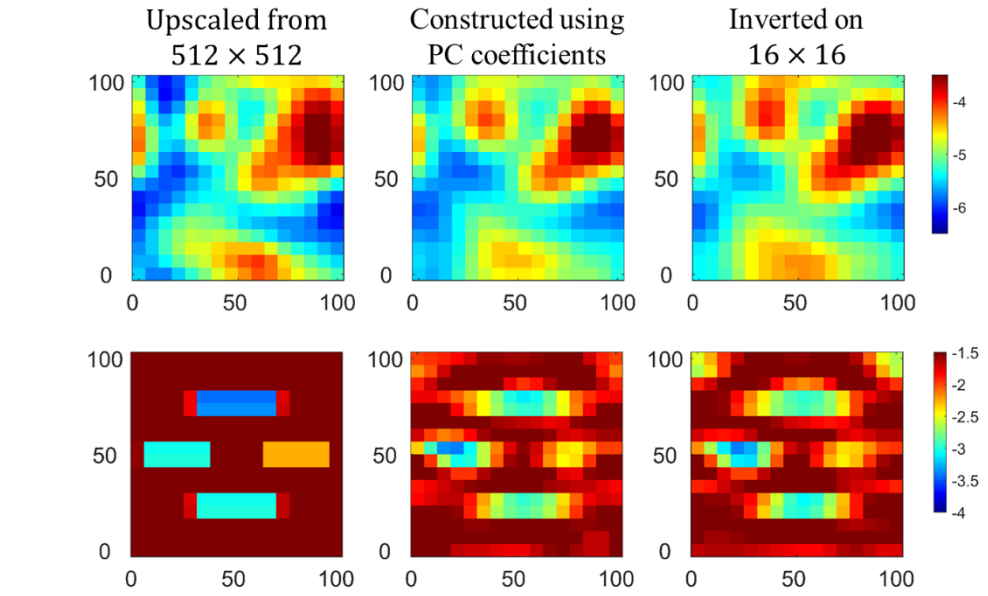

Figure 24 Left: Spatially averaged true field on 16×16 resolution. Middle: An estimate of the spatially averaged true field on 16×16, constructed by the principal component coefficients. Right: An estimate by direct inversion on the 16×16 resolution.

Results in Figure 24 also reveals a fact that the large-scale patterns are well retained by running coarse-grid forward models, which is in accordance with a finding in [*Zha et al.*, 2018]. *Zha et al.* [2018] also found that coarse grids might be more appropriate to identify large-scale hydraulic conductivity structures.

6.4 Conclusion

In this chpater, we develop an upscaling principal component geostatistical approach (UPCGA) for Bayesian inverse problems to invert high-dimensional hydraulic conductivity fields. The proposed approach is constructed on the basis of PCGA and RGA, which accelerates the traditional geostatistical inverse approach via PCA. By using the principal component coefficients as a group of latent variable, the inversion solutions on the upscaled coarse-resolution field and the fine-resolution field are connected. Thus, we are able to inverse the principal component coefficients by solely running the forward model on upscaled coarse-resolution fields, and recover the fine-resolution fields from these coefficients after the inversion process converges. Uncertainty characterization approach is also proposed accordingly based on the classical one used in PCGA and RGA, which only demands forward model runs on the coarse-resolution fields. We also proposed methods to generate conditional realizations from the approximated Gaussian posterior.

The proposed approach is tested on three synthetic numerical experiments of hydraulic tomography, and compared with PCGA. In the Gaussian random field case, UPCGA substantially reduces the total computational time of the inversion while still maintains a high accuracy in the inversion results, and yields much lower estimate variances. In the structured field case, from a practical perspective, we tested UPCGA in a

scenario with a lack of proper prior information. The structured field is modeled by a Gaussian random field during inversion. With such highly biased prior information, the robustness of UPCGA is tested and compared with PCGA. The results indicate that both algorithms are well capable of handling the biased prior information and the structured characteristics in the field. In the sequential transient pumping case, UPCGA is applied to invert transient pumping data. The results also demonstrate the effectiveness of UPCGA in handling transient data. Besides, by using transient data, which is usually featured with larger data volume, UPCGA yields much lower uncertainty estimates.

The developed approach also has the potential to take advantage of other image reduction algorithms, as long as the connections from the latent variable to the fine-resolution field and the coarse-resolution field are well established. However, the generalization of the proposed approach to other kinds of observational data, such as concentration data in tracer tests or temperature data, has not been tested. Due to the properties of transport and grid requirements of numerical stability, concentration data may be much more sensitive to the grid resolution [*Lee et al.*, 2018]. Thus, the reduction factor is not expected to be as huge as demonstrated in the above pumping cases. Considering pumping tests are much more cost-effective and common in practice, the develop approach has great potential to reduce computational efforts in aquifer characterization.

CHAPTER 7. APPLICATION OF UPCGA TO REAL WORLD DATA

7.1 Introduction

In this chapter, the proposed upscaling principal component geostatistical approach (UPCGA) is applied to investigate the hydraulic conductivity field and storage coefficient field of a deep lime stone aquifer in Xingdong coalmine site located in China. The inversion results provide more quantitative geological information of the coalmine site, and have the potential to provide guidance for coalmine operations in the future to evade incidents.

There are seven observation wells installed around the coalmine boundary to monitor the water level changes, which functions to provide early warning when incidents such as water inrush happen. These wells are shown in Figure 25 with names of D1, D3, D4, D5, D6, D9 and D11. Out of safety considerations, no monitoring well is installed in the mining area.

In order to characterize hydraulic conductivity and storage coefficient of the coalmine area, a pumping test was conducted. As shown in Figure 26, during this pumping test, in which D5 acts as the pumping well while the rest six wells act as monitoring wells, there are 14 water level readings recorded for each monitoring well. The 14 readings of each well record water levels in 7 days for every half a day. Besides the pumping test, Figure 26 also records the draw down curves for a water inrush incidents happened in W1, which can be regarded as another pumping test and complete the sequential pumping test. The data of the water inrush incident is also recorded on the same time resolution. It worth noting that during the water inrush, drawdown in D9 is much larger than those in the other

wells are. Therefore, another separated axis is adopted on the right of the figure to illustrate the range of the drawdown curve in D9.

To estimate the hydraulic conductivity and the storage coefficient, we rewrite the normal equations of UPCGA to include the storage coefficient as unknowns, and apply this joint inversion approach to the sequential pumping test data.

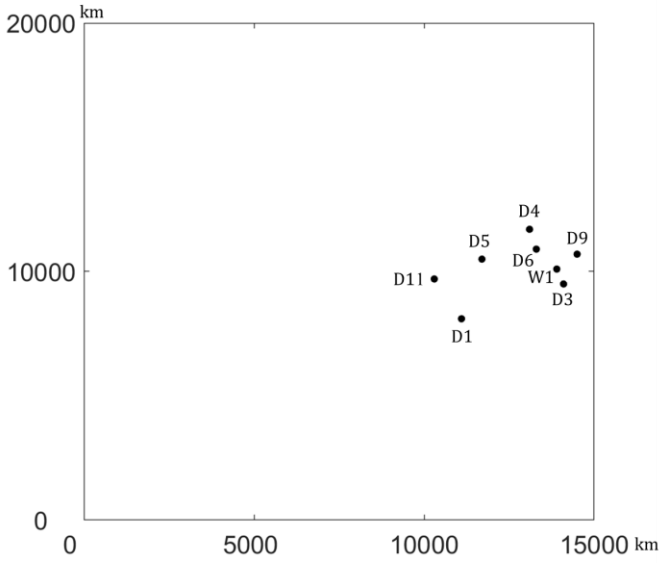

Figure 25 Setup of installed wells in the coalmine area.

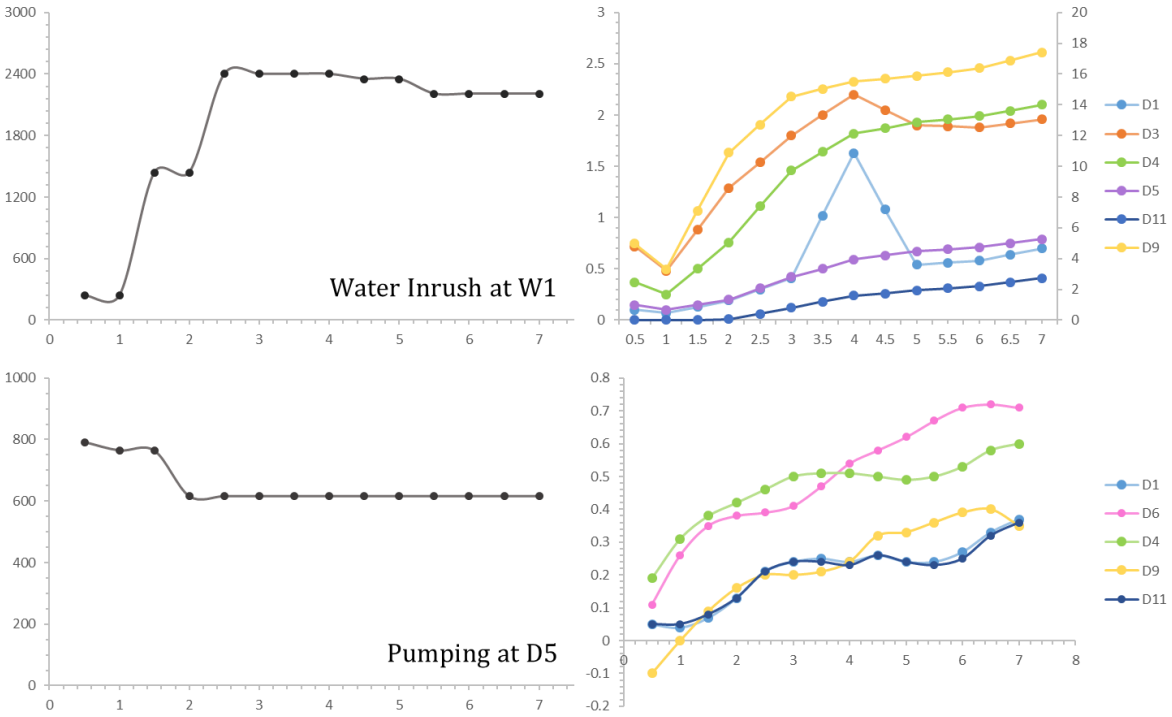

Figure 26 Draw down curves for the water inrush in the coalmine area and the pumping test.

7.2 Joint Inversion by UPCGA

In order to jointly estimate the storage coefficients and the hydraulic conductivity, the UPCGA needs to be reformulated to incorporate the variation of the storage coefficient. By applying the same geostatistical approach, a linear expansion of the natural log storage coefficient $\mathbf{q}_1$ with respect to the principal component coefficients yields

$$\mathbf{q}_1 \approx \mathbf{Y}\boldsymbol{\beta}_\mathbf{q} + \mathbf{W}_\mathrm{k}\mathbf{a}_\mathbf{q} \tag{84}$$

where $\mathbf{a}_\mathbf{q}$ is the same standard multi-Gaussian random vector as $\mathbf{a}$, $\mathbf{Y}$ reflects the deterministic trend of the field, and $\mathbf{W}_\mathrm{k} = [\mathbf{w_1}\ \mathbf{w_2}\ \mathbf{w_3}\ \ldots\ldots\ \mathbf{w_k}]$ represents the principal components of the field. By applying the upscaling matrix on $\mathbf{Y}$ and $\mathbf{W}_\mathrm{k}$, the storage coefficient field on coarse grid can be expressed as

$$\mathbf{q}_2 = \mathbf{U}\mathbf{q}_1 = \mathbf{U}\mathbf{Y}\boldsymbol{\beta}_\mathbf{q} + \mathbf{U}\mathbf{W}_\mathrm{k}\mathbf{a}_\mathbf{q} = \mathbf{Y}'\boldsymbol{\beta}_\mathbf{q} + \mathbf{W}_\mathrm{k}'\mathbf{a}_\mathbf{q} \tag{85}$$

where $\mathbf{Y}'$ and $\mathbf{W}_\mathrm{k}'$ represent the upscaled trends and principal components. Considering the effect of both hydraulic conductivity and storage coefficient, the observation equation becomes:

$$\mathbf{y} = \mathbf{f}(\mathbf{s}_2, \mathbf{q}_2) + \boldsymbol{\epsilon}_{2,2} \tag{86}$$

where $\boldsymbol{\epsilon}_{2,2}$ represents errors raised by using $\mathbf{s}_2$, $\mathbf{q}_2$ instead of $\mathbf{s}_1$, $\mathbf{q}_1$ in the transient forward model evaluations plus the original error term. For simplicity, we assume $\boldsymbol{\epsilon}_{2,2}$ to be a multi-Gaussian distribution $\mathrm{N}(\mathbf{0}, \mathbf{R}_{\mathbf{2,2}})$. In this chapter, the correlation matrix $\mathbf{R}_{\mathbf{2,2}}$ is also assumed to be $\sigma_{2,2}^2\mathbf{I}$. Accordingly, the linear system is revised to be

$$\begin{bmatrix} \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a},\boldsymbol{\beta}}\bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} + \sigma^2_{2,2}\mathbf{I} & \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a},\boldsymbol{\beta}}\bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} \\ \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a_q},\boldsymbol{\beta_q}}\bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} & \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a_q},\boldsymbol{\beta_q}}\bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} + \sigma^2_{2,2}\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_{\mathbf{q}} \\ \hat{\boldsymbol{\beta}}_{\mathbf{q}} \end{bmatrix}$$
$$= \begin{bmatrix} \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a},\boldsymbol{\beta}} \left(\mathbf{y} - \mathbf{f}(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{a}}_{\mathbf{q}}, \bar{\boldsymbol{\beta}}_{\mathbf{q}}) + \bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} \begin{bmatrix} \bar{\mathbf{a}} \\ \bar{\boldsymbol{\beta}} \end{bmatrix} + \bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} \begin{bmatrix} \bar{\mathbf{a}}_{\mathbf{q}} \\ \bar{\boldsymbol{\beta}}_{\mathbf{q}} \end{bmatrix} \right) \\ \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a_q},\boldsymbol{\beta_q}} \left(\mathbf{y} - \mathbf{f}(\bar{\mathbf{a}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{a}}_{\mathbf{q}}, \bar{\boldsymbol{\beta}}_{\mathbf{q}}) + \bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} \begin{bmatrix} \bar{\mathbf{a}} \\ \bar{\boldsymbol{\beta}} \end{bmatrix} + \bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} \begin{bmatrix} \bar{\mathbf{a}}_{\mathbf{q}} \\ \bar{\boldsymbol{\beta}}_{\mathbf{q}} \end{bmatrix} \right) \end{bmatrix} \tag{87}$$

where $\bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}}$ represents Jacobian matrix of $\mathbf{f}(\mathbf{s}_2, \mathbf{q}_2)$ with respect to $\mathbf{a}$ and $\boldsymbol{\beta}$, and $\bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}}$ represents Jacobian matrix of $\mathbf{f}(\mathbf{s}_2, \mathbf{q}_2)$ with respect to $\mathbf{a}_{\mathbf{q}}$ and $\boldsymbol{\beta}_{\mathbf{q}}$. The reformulated equation can jointly invert the hydraulic conductivity field and the storage coefficient field by only running the upscaled forward model. Between optimization iterations, $\mathbf{s}_2$ and $\mathbf{q}_2$ can be updated as

$$\mathbf{s}_2 = \mathbf{X}'\hat{\boldsymbol{\beta}} + \mathbf{V}'_{\mathrm{k}}\hat{\mathbf{a}}$$
$$\mathbf{q}_2 = \mathbf{Y}'\hat{\boldsymbol{\beta}}_{\mathbf{q}} + \mathbf{W}'_{\mathrm{k}}\hat{\mathbf{a}}_{\mathbf{q}} \tag{88}$$

which is achieved by only running the coarse-resolution forward model, and once the algorithm converges, the original fine-resolution solution $\mathbf{s}_1$, $\mathbf{q}_1$ can be recovered by Equation (76) and Equation (84) respectively.

The joint approach can also provide uncertainty quantification of the best estimates. After the iterative procedure has converged, the posterior covariance of $\mathbf{s}_1$, $\mathbf{q}_1$ can be computed as:

$$\begin{bmatrix} \mathbf{V_k}^{\mathrm{T}} \\ \mathbf{X}^{\mathrm{T}} \\ \mathbf{W_k}^{\mathrm{T}} \\ \mathbf{Y}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a},\boldsymbol{\beta}} \mathbf{R}^{-1}_{2,2} \bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} + \mathbf{I} & \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a},\boldsymbol{\beta}} \mathbf{R}^{-1}_{2,2} \bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} \\ \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a_q},\boldsymbol{\beta_q}} \mathbf{R}^{-1}_{2,2} \bar{\mathbf{H}}'_{\mathbf{a},\boldsymbol{\beta}} & \bar{\mathbf{H}}'^{\mathrm{T}}_{\mathbf{a_q},\boldsymbol{\beta_q}} \mathbf{R}^{-1}_{2,2} \bar{\mathbf{H}}'_{\mathbf{a_q},\boldsymbol{\beta_q}} + \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{V_k}^{\mathrm{T}} \\ \mathbf{X}^{\mathrm{T}} \\ \mathbf{W_k}^{\mathrm{T}} \\ \mathbf{Y}^{\mathrm{T}} \end{bmatrix} \tag{89}$$

Learning from the covariance computation in UPCGA, the above equation leverages the low-cost coarse grid forward model to produce variance maps in short computational time. As the off-diagonal elements of this matrix encodes cross-covariance matrices between $\mathbf{s}_1$ and $\mathbf{q}_1$, which are of no interest in the uncertainty quantification, the computational cost only concentrates on finding the diagonal elements, which can be efficiently computed as what has been demonstrated in the former chapter.

7.3 Real World Data Inversion

The inversion of real world data always requires a prior guess of the field characteristics, which, in the geostatistical background, are governed by structural parameters of the field. The initialization of the geostatistical parameters are shown in Table 10, which reflects our prior belief in the hydraulic conductivity field and the storage coefficient field. Other values of the parameters are also explored in data inversion; however, the inversion of data does not produce significant differences across different setups. This experimental result also corroborates the statement that certain dataset supports more than one prior belief. A group of similar geostatistical parameters can provide similarly accurate inversion result, which is also in accordance with what we found in 6.3.2 and [*Liu and Kitanidis*, 2011].

Table 10 Initialization of parameters in the real world data inversion.

Geostatistical Properties	
Domain Scale	$15000\text{m} \times 20000\text{m}$
Field Type	Logarithmic Field
Resolution	750×1000
Covariance Model	Exponential model
Mean	$\mu = 1.2\text{m/s}$ $\mu = -9.0$
Variance	$\sigma^2 = 1.5\text{m}^2/\text{s}^2$ $\sigma^2 = 1.5$
Correlation length	$\text{lx} = 1000\text{m}, \text{ly} = 1000\text{m}$
Governing Equation	
$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)\left(K\left(\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\right)\right) = Q + S\frac{\partial h}{\partial t}$	
North	$h = 950\text{m}$
South	$h = 950\text{m}$
West	$h = 950\text{m}$
East	Impermeable

The inversion is solved by UPCGA that upscales the fields from the original resolution to 75×100. The same upscaling scheme is introduced to both fields. The MAP estimates of the logarithmic hydraulic conductivity and the logarithmic storage coefficient are shown in Figure 27, with pumping wells denoted as white dots and monitoring wells denoted as black dots.

The result indicates that the studied area possesses highly heterogeneous behaviors in both conductivity field and storage coefficient field. The magnitude of the hydraulic conductivity **K** ranges from 10^{-2} to 10^{2}, while that of the storage coefficient **S** ranges from 10^{-4} to 10^{-6}. This also helps to explain the contrasting drawdown curves from different monitoring wells. Besides, the result also reveals the spatial pattern of conductivity and storage coefficient in the studied area.

The inversion of field data uncovers two major extreme-conductivity regions in the studied area. One low conductivity area residing on the east, impermeable boundary that embraces D4, D6, D9 and W1, and one high conductivity area locating to the south of the well network that covers D1 and D11. Storage coefficient field owns a 'clustered' pattern, which is characterized by one high storage region and one low storage region, while the rest of it is almost uniform around the mean value. The extremely low storage coefficient distribution on the east boundary prompts the drawdowns in D4, D6 and D9 to hit the plateau within shorter pumping duration compared with D1, D5 or D11 that locate in the high storage region.

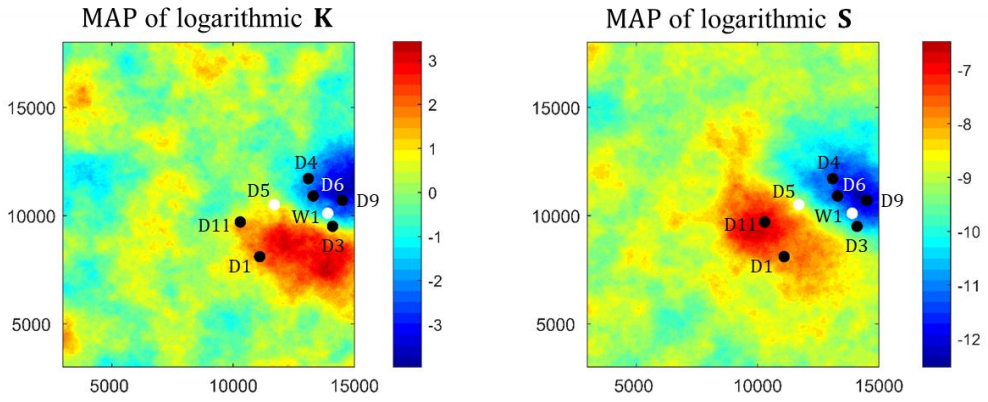

Figure 27 Best estimates of the logarithmic hydraulic conductivity field and the logarithmic storage field.

The uncertainty quantifications of conductivity and storage coefficient are computed via Equation (89), and the variance of each grid point, which are represented by the diagonal elements in Equation (89), are plotted in Figure 28. Both variance maps demonstrate similar spatial patterns: a much lower uncertainty in the region where the well network is constructed and almost a uniform distribution of medium uncertainty otherwise. The low uncertainty region carries variance value between 0.1-0.15, which is around 10 percent of the total variance of the logarithmic conductivity and 30 percent of the total variance of the logarithmic storage coefficient. The quantitative analysis indicates that even though transient pumping data enables the inversion of storage coefficient, it contains more confidence in respect of inverting the conductivity.

It is also worth noting that the spatial pattern shown in the variance maps is also consistent with the conclusion we made in CHAPTER 5 and CHAPTER 6. Measurements reveal local details of where they are taken while at the same time increases certainty of estimate in this region.

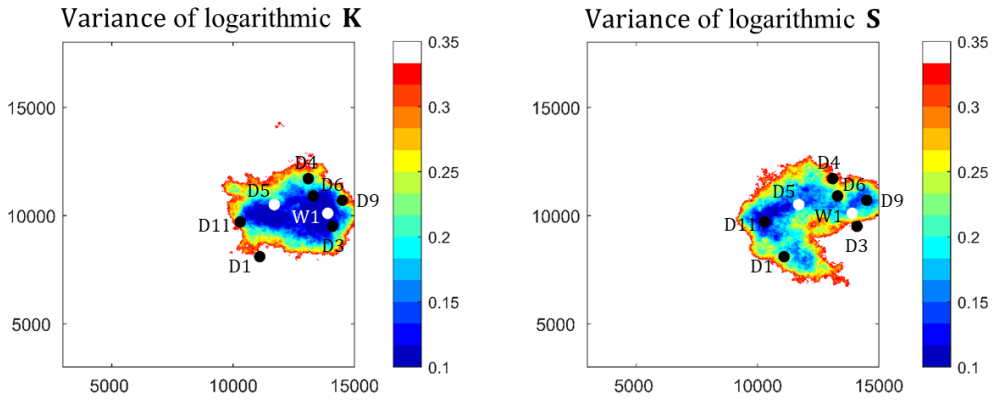

Figure 28 Variance maps of the logarithmic hydraulic conductivity field and the logarithmic storage field.

The performance of UPCGA in inverting field data is also investigated from the perspective of reproducing the measurement. Though inverted jointly, the two datasets, pumping test data and water inrush data, are compared separately in Figure 29 and Figure 30.

The results in Figure 29 indicate that the proposed approach successfully reproduces the drawdown curves of the pumping test, which is conducted under a normal pumping rate. However, the measurement is not as well reproduced in the water inrush incident as in the pumping test, especially in D3, D5 and D11. The reason can be two-fold. First, unlike a pumping test that is conducted by researchers and engineers with deliberately designed setup, water inrush incident happens more abruptly with an extreme flow rate, which potentially incurs unexpected hydraulic behaviors such as change of water inrush channels during the incident [*Mao et al.*, 2018] or highly fluctuated flow rate. Second, the pumping of water from the coalmine does not start instantaneously after the water inrush happens, and the pumping rate may not accurately reflect the real inrush flow rate in the coalmine. However, we shall also notice that UPCGA is strongly challenged by reproducing extreme drawdown curve, which could be illustrated by data points in D9 in Figure 30.

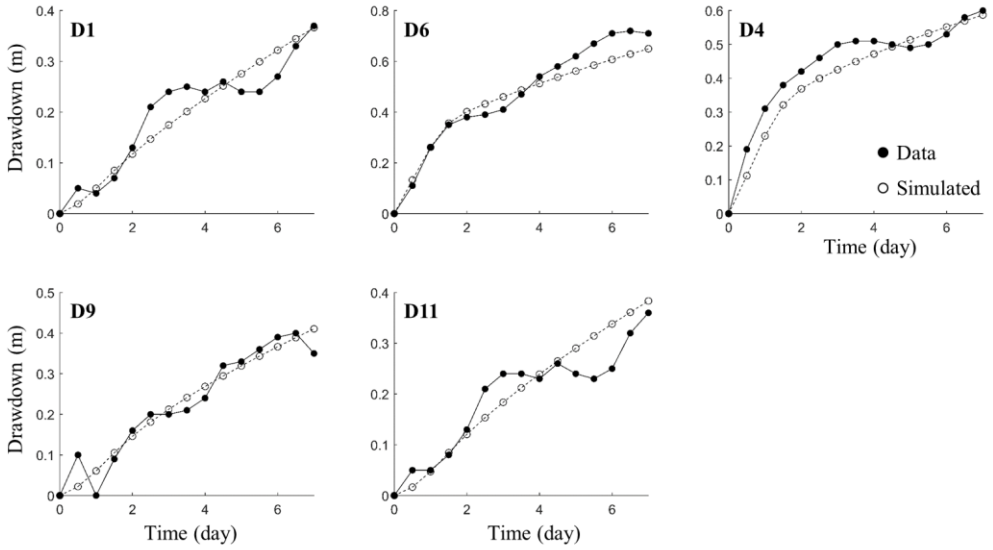

Figure 29 Reproduction of measurement for pumping test.

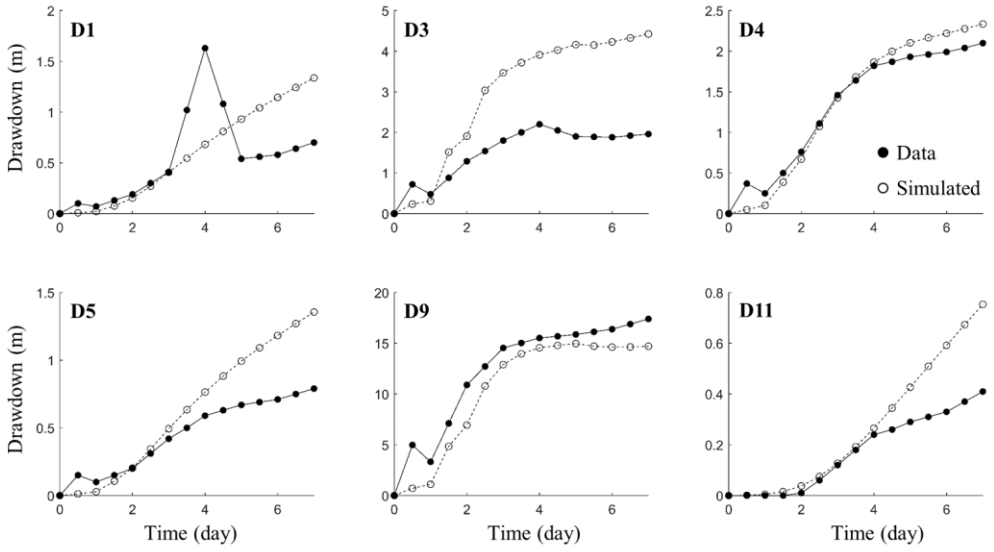

Figure 30 Reproduction of measurement for water inrush incident.

7.4 Conclusion

In this chapter, we investigate the performance of the proposed UPCGA inverse algorithm with a real-world application: a large-scale inverse problem with transient pumping data. The studied domain stretches 15km long and 20km high with the discretized resolution 750×1000. To alleviate the prohibitive computational overhead, the original resolution is upscaled to 75×100 in the upscaling stage.

The inversion result reveals structural formations having high or low hydraulic conductivity and storage coefficients. The uncovered spatial patterns are aligned with what has been found by other researchers [*Mao et al.*, 2018], and contribute to future study of this coalmine field. At the same time, UPCGA also captures the dominant trend of the measured drawdown curve, especially for the drawdown data from the pumping test. Even though a validation process can help to substantiate our finding, due to a lack of datasets, we decided not to employ it. For future study, if there are more data chunks available, our algorithm is well capable of assimilating these new data to strengthen the delineation of geological formations in this area.

CHAPTER 8. CONCLUSION

In this thesis, we develop new inverse modeling algorithms in a Bayesian framework and demonstrate their applications in estimating large-dimensional hydrogeological parameter fields. Specifically, we develop new, efficient and accurate algorithms for both stochastic sampling approach, i.e., MCMC, and objective approach, i.e., Bayesian geostatistical approach.

In CHAPTER 3, based on the stochastic Newton approach, a quasi-Newton approach is proposed and assimilated into the sampling framework to reduce the number of forward model evaluations. The proposed quasi-Newton approach is called Broyden method, which is designed for Jacobian approximation in optimization problems at the first place. The results of numerical experiments indicate that Broyden algorithm substantially cut down the total computational cost while maintains a high accuracy of inversion results.

In CHAPTER 4, we construct an upscaling approach for large-scale hydraulic tomography problems, in which a group of latent variable is derived to encode the information across solutions on different resolutions. Therefore, the forward model can be evaluated with minor computational effort on the upscaled field. The upscaling approach is combined with the sampling framework investigated in CHAPTER 3 to provide further computational advantage to handle the inversion of hydraulic data. The performance of this approach is scrutinized from the perspective of computational budget and inversion result quality. The results indicate that the proposed approach significantly accelerates the process of solution-finding while maintains a high accuracy in the final estimate.

In CHAPTER 5 and CHAPTER 6, we reformulate the geostatistical approach that transforms the unknown random variable to be the principal component coefficients, and incorporates the upscaling approach into the geostatistical approach to develop an upscaling principal component geostatistical approach (UPCGA). The reformulated geostatistical approach (RGA) possesses a more intuitive way of computing Jacobian matrices, and yields a scalable normal equation system that particularly benefits inverse problems with a large volume of measurements. UPCGA takes advantage of RGA and recognizes the principal component coefficients as latent variable to connect the coarse-grid solution and the fine-grid solution. Several numerical experiments are used to verify the proposed algorithms. The experimental results demonstrate the capability of RGA and UPCGA in handling high-resolution Bayesian inverse problems, especially in the UPCGA case, the computational cost reduction is substantial.

In CHAPTER 7, the performance of UPCGA is investigated by a field data case. The dataset is collected by conducting a pumping test and a water inrush incident in a coalmine site. The proposed approach inverts the field with alleviated computational cost, while simultaneously identifies the critical regions of conductivity and storage coefficient in the studied area. The uncovered spatial patterns are aligned with what has been found by other researchers [*Mao et al.,* 2018], and contribute to future exploration operation of this coalmine field. However, we also noticed that the reproduction of extreme drawdown data points is challenging for UPCGA.

It is worth noting that MCMC algorithms are usually applied for the cases when future computation based on the generated realizations are expected, whereas GA focuses more on providing a single fast and reliable estimate of the studied field.

As far as we know, the newly derived Bayesian inverse algorithms also have the potential to be extended to other inverse applications such as tracer data inversion, seismic wave inversion or temperature data inversion. A decomposition and recombination of the proposed approaches can potentially stimulate new algorithms that absorb the advantages of both. Besides, the future of Bayesian inverse modeling may also become intertwined with the rapid development of deep learning and variational Bayesian, which lends sophisticated characterization tool and more efficient computational tool to current Bayesian inverse modeling community.

APPENDIX.

REFERENCES

Aharon, M., Elad, M., & Bruckstein, A. (2006). $ rm k $-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing, 54*(11), 4311-4322.

Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. (2013). Large-scale stochastic linear inversion using hierarchical matrices. *Computational Geosciences, 17*(6), 913-927.

Bates, B. C., & Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research, 37*(4), 937-947.

Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation, 19*(92), 577-593.

Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., & Wilcox, L. C. (2012). *Extreme-scale UQ for Bayesian inverse problems governed by PDEs.* Paper presented at the Proceedings of the international conference on high performance computing, networking, storage and analysis.

Bui-Thanh, T., & Girolami, M. (2014). Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo. *Inverse Problems, 30*(11), 114014.

Caers, J., & Zhang, T. (2004). Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models.

Cardiff, M., Barrash, W., Kitanidis, P., Malama, B., Revil, A., Straface, S., & Rizzo, E. (2009). A potential-based inversion of unconfined steady-state hydraulic tomography. *Groundwater, 47*(2), 259-270.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., & Slooten, L. J. (2005). Inverse problem in hydrogeology. *Hydrogeology journal, 13*(1), 206-222.

Carrera, J., & Neuman, S. P. (1986). Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research, 22*(2), 199-210.

Chen, J., Kemna, A., & Hubbard, S. S. (2008). A comparison between Gauss-Newton and Markov-chain Monte Carlo–based methods for inverting spectral induced-polarization data for Cole-Cole parameters. *Geophysics, 73*(6), F247-F259.

Chen, P. (2011). Hessian matrix vs. Gauss–Newton hessian matrix. *SIAM Journal on Numerical Analysis, 49*(4), 1417-1435.

Cirpka, O. A., Fienen, M. N., Hofer, M., Hoehn, E., Tessarini, A., Kipfer, R., & Kitanidis, P. K. (2007). Analyzing bank filtration by deconvoluting time series of electric conductivity. *Groundwater, 45*(3), 318-328.

Cui, T., Fox, C., & O'sullivan, M. (2011). Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. *Water Resources Research, 47*(10).

Cui, T., Martin, J., Marzouk, Y. M., Solonen, A., & Spantini, A. (2014). Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems, 30*(11), 114015.

Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X., & Mallick, B. (2005). An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resources Research, 41*(12).

Elsheikh, A. H., Wheeler, M. F., & Hoteit, I. (2014). Hybrid nested sampling algorithm for Bayesian model selection applied to inverse subsurface flow problems. *Journal of Computational Physics, 258*, 319-337.

Fienen, M. N., Clemo, T., & Kitanidis, P. K. (2008). An interactive Bayesian geostatistical inverse protocol for hydraulic tomography. *Water Resources Research, 44*(12).

Fienen, M. N., Luo, J., & Kitanidis, P. K. (2006). A Bayesian geostatistical transfer function approach to tracer test analysis. *Water Resources Research, 42*(7).

Fu, J., & Gómez-Hernández, J. J. (2009). Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method. *Journal of Hydrology, 364*(3), 328-341.

Gavalas, G., Shah, P., & Seinfeld, J. H. (1976). Reservoir history matching by Bayesian estimation. *Society of Petroleum Engineers Journal, 16*(06), 337-350.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Geman, S., & Geman, D. (1984a). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 721-741.

Geman, S., & Geman, D. (1984b). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721-741.

Girolami, M., & Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(2), 123-214.

Gottlieb, J., & Dietrich, P. (1995). Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Problems, 11*(2), 353.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97-109.

Hoeksema, R. J., & Kitanidis, P. K. (1984). An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *20*(7), 1003-1020.

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(1), 1593-1623.

Hu, W., Abubakar, A., Habashy, T., & Liu, J. (2011). Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion. *Geophysical Prospecting, 59*(3), 477-491.

Jafarpour, B., & McLaughlin, D. B. (2009). Reservoir characterization with the discrete cosine transform. *SPE Journal, 14*(01), 182-201.

Ju, L., Zhang, J., Meng, L., Wu, L., & Zeng, L. (2018). An adaptive Gaussian process-based iterative ensemble smoother for data assimilation. *Advances in Water Resources, 115*, 125-135.

Kang, P. K., Lee, J., Fu, X., Lee, S., Kitanidis, P. K., & Juanes, R. (2017). Improved characterization of heterogeneous permeability in saline aquifers from transient pressure data during freshwater injection. *Water Resources Research, 53*(5), 4444-4458.

Khaninezhad, M. M., & Jafarpour, B. (2014). Sparse Randomized Maximum Likelihood (SpRML) for subsurface flow model calibration and uncertainty quantification. *Advances in Water Resources, 69*, 23-37.

Kitanidis, & Vomvoris, E. G. (1983a). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *19*(3), 677-690.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research, 22*(4), 499-507.

Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research, 31*(10), 2411-2419.

Kitanidis, P. K. (1996). On the geostatistical approach to the inverse problem. *Advances in Water Resources, 19*(6), 333-342.

Kitanidis, P. K. (1997). *Introduction to geostatistics: applications in hydrogeology*: Cambridge University Press.

Kitanidis, P. K., & Lee, J. (2014). Principal Component Geostatistical Approach for large-dimensional inverse problems. *Water Resources Research, 50*(7), 5428-5443.

Kitanidis, P. K., & Vomvoris, E. G. (1983b). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research, 19*(3), 677-690.

Klein, O., Cirpka, O. A., Bastian, P., & Ippisch, O. (2017). Efficient geostatistical inversion of transient groundwater flow using preconditioned nonlinear conjugate gradients. *Advances in Water Resources, 102*, 161-177.

Klute, A., & Dinauer, R. C. (1986a). Physical and mineralogical methods. *Planning, 8*, 79.

Klute, A., & Dinauer, R. C. J. P. (1986b). Physical and mineralogical methods. *8*, 79.

Köpke, C., Irving, J., & Roubinet, D. (2019). Stochastic inversion for soil hydraulic parameters in the presence of model error: An example involving ground-penetrating radar monitoring of infiltration. *Journal of Hydrology, 569*, 829-843.

Kruseman, G. P., De Ridder, N. A., & Verweij, J. M. (1970). *Analysis and evaluation of pumping test data* (Vol. 11): International institute for land reclamation and improvement The Netherlands.

Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network. *Water Resources Research, 54*(1), 381-406.

Lanczos, C. (1950). *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*: United States Governm. Press Office Los Angeles, CA.

Law, K. J. (2014). Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics, 262*, 127-138.

Lee, J., & Kitanidis, P. K. (2014). Large-scale hydraulic tomography and joint inversion of head and tracer data using the Principal Component Geostatistical Approach (PCGA). *Water Resources Research, 50*(7), 5410-5427.

Lee, J., Kokkinaki, A., & Kitanidis, P. K. (2018). Fast large-scale joint inversion for deep aquifer characterization using pressure and heat tracer measurements. *Transport in Porous Media, 123*(3), 533-543.

Lee, J., Yoon, H., Kitanidis, P. K., Werth, C. J., & Valocchi, A. J. (2016). Scalable subsurface inverse modeling of huge data sets with an application to tracer

concentration breakthrough data from magnetic resonance imaging. *Water Resources Research, 52*(7), 5213-5231.

Li, W., & Cirpka, O. A. (2006). Efficient geostatistical inverse methods for structured and unstructured grids. *Water Resources Research, 42*(6).

Li, W., Englert, A., Cirpka, O. A., Vanderborght, J., & Vereecken, H. (2007). Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests. *Water Resources Research, 43*(4).

Li, W., Englert, A., Cirpka, O. A., & Vereecken, H. (2008). Three-dimensional geostatistical inversion of flowmeter and pumping test data. *Groundwater, 46*(2), 193-201.

Li, W., Nowak, W., & Cirpka, O. A. (2005). Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. *Water Resources Research, 41*(8).

Liao, Q., Zhang, D., & Tchelepi, H. (2017). A two-stage adaptive stochastic collocation method on nested sparse grids for multiphase flow in randomly heterogeneous porous media. *Journal of Computational Physics, 330*, 828-845.

Liao, Z., & Cirpka, O. A. (2011). Shape-free inference of hyporheic traveltime distributions from synthetic conservative and "smart" tracer tests in streams. *Water Resources Research, 47*(7).

Lin, Y., Le, E. B., O'Malley, D., Vesselinov, V. V., & Bui-Thanh, T. J. W. R. R. (2017). Large-scale inverse model analyses employing fast randomized data reduction. *Water Resources Research, 53*(8), 6784-6801.

Liu, S., Yeh, T. C. J., & Gardiner, R. (2002). Effectiveness of hydraulic tomography: Sandbox experiments. *Water Resources Research, 38*(4), 5-1-5-9.

Liu, X., Illman, W., Craig, A., Zhu, J., & Yeh, T. C. J. W. R. R. (2007). Laboratory sandbox validation of transient hydraulic tomography. *43*(5).

Liu, X., & Kitanidis, P. (2011). Large-scale inverse modeling with an application in hydraulic tomography. *Water Resources Research, 47*(2).

Liu, Y. (2006). Using the Snesim program for multiple-point statistical simulation. *Computers & geosciences, 32*(10), 1544-1563.

Luo, J., & Cirpka, O. A. (2011). How well do mean breakthrough curves predict mixing-controlled reactive transport? *Water Resources Research, 47*(2), W02520.

Mao, D., Liu, Z., Wang, W., Li, S., Gao, Y., Xu, Z., & Zhang, C. (2018). An application of hydraulic tomography to a deep coal mine: combining traditional pumping tests with water inrush incidents. *Journal of Hydrology, 567*, 1-11.

Martin, J., Wilcox, L. C., Burstedde, C., & Ghattas, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing, 34*(3), A1460-A1487.

Matheron, G. (1963). Principles of geostatistics. *Economic geology, 58*(8), 1246-1266.

McLaughlin, D., & Townley, L. R. (1996). A reassessment of the groundwater inverse problem. *Water Resources Research, 32*(5), 1131-1161.

Michalak, A. M., & Kitanidis, P. K. (2003). A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification. *Water Resources Research, 39*(2).

Minasny, B., & McBratney, A. B. (2005). The Matérn function as a general model for soil variograms. *Geoderma, 128*(3-4), 192-207.

Mondal, A., Efendiev, Y., Mallick, B., & Datta-Gupta, A. (2010). Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods. *Advances in Water Resources, 33*(3), 241-256.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo, 2*, 113-162.

Neuman, S. P. (1980). A statistical approach to the inverse problem of aquifer hydrology: 3. Improved solution method and added perspective. *Water Resources Research, 16*(2), 331-346.

Nowak, W., & Cirpka, O. A. (2004). A modified Levenberg–Marquardt algorithm for quasi-linear geostatistical inversing. *Advances in Water Resources, 27*(7), 737-750.

Nowak, W., & Cirpka, O. A. (2006a). Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data. *42*(8).

Nowak, W., & Cirpka, O. A. (2006b). Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data. *Water Resources Research, 42*(8).

Nowak, W., Tenkleve, S., & Cirpka, O. A. (2003). Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geology, 35*(1), 53-66.

Pollock, D., & Cirpka, O. A. (2008). Temporal moments in geoelectrical monitoring of salt tracer experiments. *Water Resources Research, 44*(12).

Reynolds, A. C., Zafari, M., & Li, G. (2006). *Iterative forms of the ensemble Kalman filter.* Paper presented at the ECMOR X-10th European Conference on the Mathematics of Oil Recovery.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science, 16*(4), 351-367.

Roberts, G. O., & Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 341-363.

Rubin, Y., Chen, X., Murakami, H., & Hahn, M. (2010). A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. *Water Resources Research, 46*(10).

Saibaba, A. K., Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. F. (2012). Application of hierarchical matrices to linear inverse problems in geostatistics. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles, 67*(5), 857-875.

Saley, A. D., Jardani, A., Ahmed, A. S., Raphael, A., & Dupont, J. (2016). Hamiltonian Monte Carlo algorithm for the characterization of hydraulic conductivity from the heat tracing data. *Advances in Water Resources, 97*, 120-129.

Sarma, P., Durlofsky, L. J., & Aziz, K. (2008). Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences, 40*(1), 3-32.

Schölkopf, B., Smola, A., & Müller, K.-R. (1997). *Kernel principal component analysis.* Paper presented at the International Conference on Artificial Neural Networks.

Tompkins, M. J., Fernandez-Martinez, J. L., Mukerji, T., & Alumbaugh, D. L. (2010). Efficient nonlinear inverse uncertainty computation using parameter reduction, constraint mapping, and very sparse posterior sampling: Journal of Geophysical Research. *SCRF annual report*.

Tonkin, M., & Doherty, J. (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research, 41*(10).

Tonkin, M., & Doherty, J. (2009). Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resources Research, 45*(12).

Tsai, F. T. C., & Li, X. (2008). Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window. *Water Resources Research, 44*(9).

Vrugt, J. A., Stauffer, P. H., Wöhling, T., Robinson, B. A., & Vesselinov, V. V. (2008). Inverse modeling of subsurface flow and transport properties: A review with new developments. *Vadose Zone Journal, 7*(2), 843-864.

Vrugt, J. A., & Ter Braak, C. (2011). DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences, 15*(12).

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation, 10*(3), 273-290.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems, 2*(1-3), 37-52.

Woodbury, A. D., & Rubin, Y. (2000). A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site. *Water Resources Research, 36*(1), 159-171.

Yeh, T. C. J., Gutjahr, A. L., & Jin, M. (1995). An iterative cokriging-like technique for ground-water flow modeling. *Groundwater, 33*(1), 33-41.

Yeh, T. C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research, 32*(1), 85-92.

Yeh, T. C. J., & Liu, S. (2000). Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research, 36*(8), 2095-2105.

Yeh, W. W. G. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research, 22*(2), 95-108.

Yin, D., & Illman, W. A. (2009). Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study. *Water Resources Research, 45*(1).

Yoon, H., & McKenna, S. A. (2012). Highly parameterized inverse estimation of hydraulic conductivity and porosity in a three-dimensional, heterogeneous transport experiment. *Water Resources Research, 48*(10).

Zeng, L., Shi, L., Zhang, D., & Wu, L. (2012). A sparse grid based Bayesian method for contaminant source identification. *Advances in Water Resources, 37*, 1-9.

Zha, Y., Yeh, T. C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. J. W. R. R. (2018). A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. *Water Resources Research, 54*(3), 1616-1632.

Zhang, J., Lin, G., Li, W., Wu, L., & Zeng, L. J. W. r. r. (2018). An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions. *54*(3), 1716-1733.

Zhang, Z., Jafarpour, B., & Li, L. (2014). Inference of permeability heterogeneity from joint inversion of transient flow and temperature data. *Water Resources Research, 50*(6), 4710-4725.

Zhao, Y., Rathore, S. S., Liu, M., & Luo, J. (2018). Joint Bayesian Inversion for Analyzing Conservative and Reactive Breakthrough Curves. *Journal of Hydrology*.

Zhou, H., Li, L., Franssen, H.-J. H., & Gómez-Hernández, J. J. J. M. G. (2012). Pattern recognition in a bimodal aquifer using the normal-score ensemble Kalman filter. *44*(2), 169-185.

Zhu, J., & Yeh, T.-C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research, 41*(7).

Zhu, J., & Yeh, T. C. J. (2006). Analysis of hydraulic tomography using temporal moments of drawdown recovery data. *Water Resources Research, 42*(2).

Zimmerman, D., Marsily, G. d., Gotway, C. A., Marietta, M. G., Axness, C. L., Beauheim, R. L., et al. (1998). A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. *Water Resources Research, 34*(6), 1373-1413.